Fundamentals of Aerospace: (draft) course notes (engineering part) and tutorial sheets

T. Rendall

March 17, 2023

Contents

This document in a work-in-progress, designed to support the material in the lecture slides. Please use it as a second source, to help clarify what you may not have understood from the lectures themselves, but keep the lectures/slides as the primary source.

The tutorial sheet questions at the end are designed to be representative of final assessment type questions, although the hardest ones are more tricky than would be seen on the assessment generally.

NB - disclaimer - some of the examples here are from military applications. I'm just going to state the obvious; I am here to teach you how things work, not recommend or suggest their use, and nor would I ever intend to. Many creative solutions have been developed for military purposes and it's important you understand what they are, because their applications are diverse. Satellite navigation guides weapons to their targets, but also rescuers to a sinking ship. Proportional navigation guides missiles to targets but also helps prevent collisions at sea. The only thing that is really certain is you're unlikely to improve anything unless you first understand it. You can always integrate twice to find position, but your destination is up to you.

1 Non-dimensional numbers

We live in a world where measuring things is an obsession. This is human nature, but it does lead to a preoccupation with the units themselves. If we measure a length, what we're actually doing is recording is size *relative* to the reference of the metre unit. If we measure our height, what matters is not the measurement in metres, but the ratio between our height and the height of the doorway, because that is what determines whether we will bump our head or not. So, the number that matters (the 'non-dimensional bumping head number') is $B = \frac{h}{d}$, where h is your height and d the doorway height. Knowing h alone obviously tells you *nothing* about the outcome for your forehead.

That's a trivial example of course, where we're only working with length. If we include the added thrill of mass and time alongside length, we can reach a much more useful level of insight. Eventually, this train of thought will lead us to the nondimensional force coefficients used in aerodynamics, but first it's helpful to think about it more generally using some slightly more unusual examples.

1.1 The sinking ship problem

Consider a 'simple' scaling problem: you know how long it takes a model ship to sink in calm waters t_1, and now you want to know how that would scale up to the full sized ship to find t_2. The model ship has a length of l_1 and the full sized ship is l_2. Spoiler alert, it's not a linear relationship, so $t_2 \neq t_1\frac{l_2}{l_1}$. Engineering is a straightforward subject, but not a linear one!

Variables of interest include

- mass m, kg
- gravity g, ms^{-2}
- length l, m
- time t, s

- density ρ, kgm^{-3}

That's 5 variables and 3 dimensions involved (mass, length, time), which suggests 2 non-dimensional numbers may be built (because each dimension gives a relationship that links that dimension, there can only be 2 parameters that are genuinely free).

There are different ways to build the non-dimensional numbers, but it doesn't matter too much, because they are all equivalent in what they allow in terms of analysis. It's perfectly reasonable to do them by hand, just constructing them until the dimensions vanish. So let's define two numbers as

$$P_{sink} = \frac{mgt^2}{\rho_w l^4} = \frac{kgms^{-2}s^2}{kgm^{-3}m^4} = [0] = \frac{\text{gravitational force on ship}}{\text{inertial force on water}} \quad (1)$$

$$P_{float} = \frac{mg}{\rho_w g l^3} = \frac{kgms^{-2}}{kgm^{-3}ms^{-2}m^3} = [0] = \frac{\text{weight of ship}}{\text{weight of water}} \quad (2)$$

Both numbers represent a ratio of important forces in the system. Keeping P_{float} unchanged between model and full scale means that both the model and full scale vessel should float with the same non-dimensional distance of their hull beneath the water (which seems pretty obvious, really!) - see figure 1.

Keeping P_{float} fixed gives a redefined $P_{sink} = \frac{gt^2}{l}$ (because P_{float} appears in P_{sink}). The interesting point is now that $P_{sink_1} = P_{sink_2}$ - there is only a single value of P_{sink} for both the model and full-scale ship. Therefore

$$t_2 = t_1\sqrt{\frac{l_2}{l_1}} \quad (3)$$

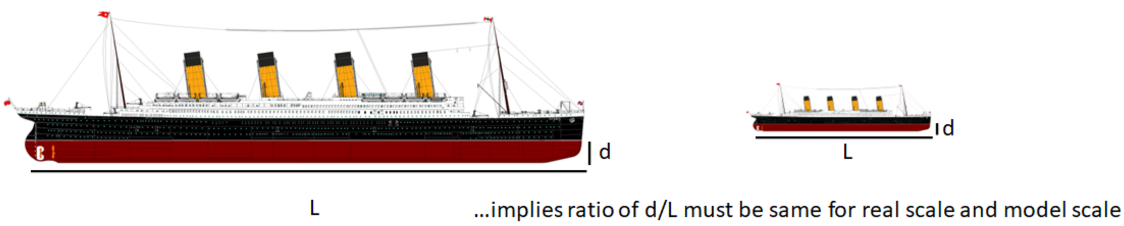


Figure 1: Non-dimensional P_{float} for a ship

So, the sinking time (or any time, for that matter) scales with the square root of length.

Being the suspicious sort, I thought it would be sensible to double check this with an experiment. Figure 2 shows three tins, all cut so that the height/diameter ratio is the same in each case. Each tin is also ballasted with weights, so that when it floats (before sinking) they all float with the same fraction of the diameter below the water surface. This means P_{float} is the same in all cases.

Figure 2 also shows the recorded sinking times, plotted as a ratio relative to the smallest tin. This is a good example of a limit to dimensional analysis - it doesn't predict P_{sink} for a particular tin (life isn't full of free lunches, afterall), but only how it varies depending on length scale. If you wanted to know P_{sink} theoretically, without doing any experiments, you would need to resort to fluid mechanics analysis of some sort (eg. similar to Toricelli's law).

As you can see in figure 2, the experimental results are quite close to the square root relationship. In any case, it is *certainly not linear*. As backyard experiments go, this is pretty good confirmation.

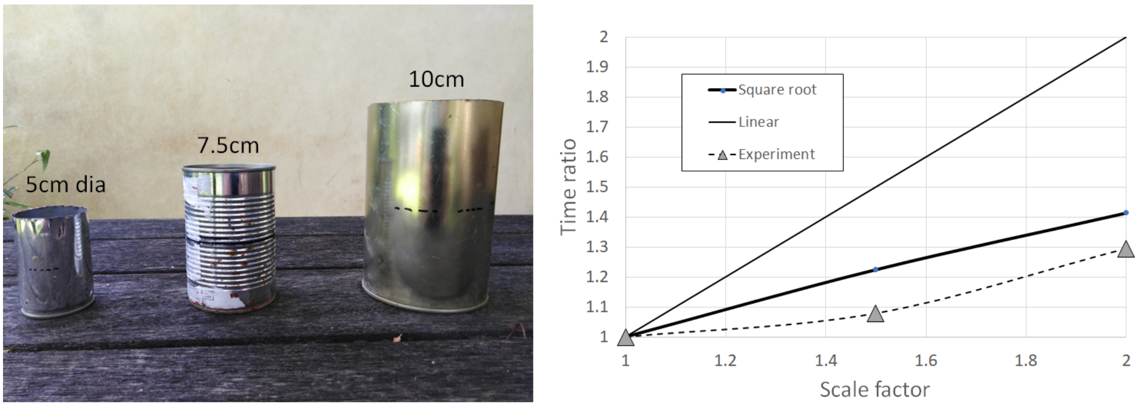


Figure 2: Sinking times of a scaled tin

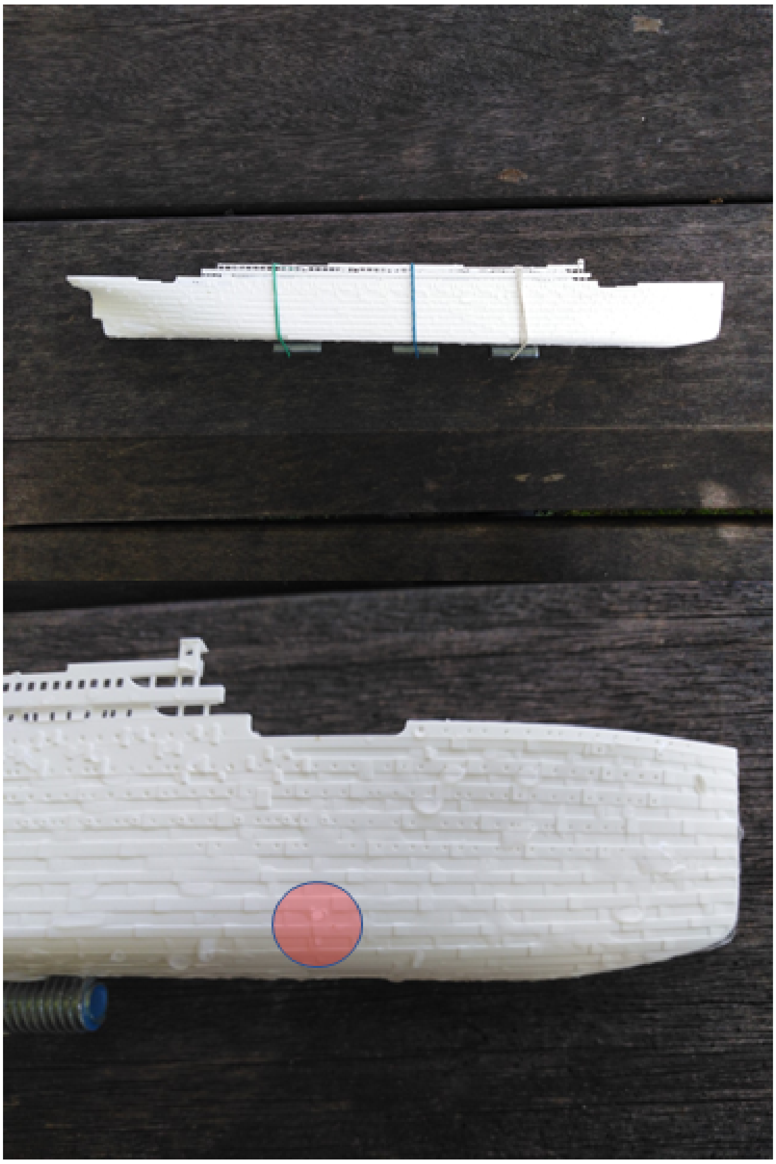


Figure 3: Scaled Titanic model

If you're thinking there must be more to this problem - yes, of course there is. In the analysis, we ignored viscosity μ of the water as a variable. If we had retained it, we would have discovered the Reynolds number Re as another dimensionless number, and technically this would have to be constant for all the tins in order to achieve the square root relationship. The fact that it isn't constant between them all is one reason why the results may not be linear (or I may be a poor user of stopwatches). The same argument holds for surface tension effects. The moral of the story is - the dimensional analysis only holds for the variables you choose, and if you leave some out, then you are implicitly assuming those dimensionless numbers are constant. If they are not, then you may not see the expected behaviour.

Of course, engineering is all about prudent assumptions, so this realisation is expected. Engineers make the *right* assumptions to work out what they want to know. We do dot spend our spend complaining about the impossibility of resolving all physical behaviour exactly (that's a different department).

Now, let's apply what we know to the most famous sinking, of RMS *Titanic*. Purchasing a 1:700 scale model, and with the use of a large baking tin, we can resolve this problem. The model is shown in figure 3, ballasted to float at the right depth, and with a 1.5mm diameter hole drilled in the forward hull. The m^2 area of the hole on the real ship was about the size of a doorway in total, as estimated by the later investigation, so this is a scaled area. Of course, the real damage was distributed along the hull, but this cannot be represented at this scale.

The model sinking time was found to be 7.17 minutes, giving a full-scale sinking time of

$$7.17 \times \sqrt{700} = 190\text{minutes} \tag{4}$$

and compares reasonably well with the real reported time of 160 minutes. But wait...I hear your scepticism brewing. What sources of error could there be?

- Viscous effects ignored - no Reynolds number scaling
- Surface tension ignores - so no Weber number effects. In fact, I had to 'wet' the hole to get the model to sink, indicating this is potentially a substantial factor
- Ship did not fracture - although, this is probably a more minor source of error, as fracture occured quite late in the process
- Ship geometry not exact - no watertight compartments, and hull damage only approximate

So, of course, this is not a highly accurate result. If you wanted a really good result, you would need to build a model several metres long at least, and persuading the faculty to fund this would be beyond the scope of AVDASI 1.

In case you are wondering, this is a process much used today. The sinking of MV *Estonia* in 1994 was followed by numerous theoretical and experimental investigations, with the experiments all using Froude number scaling (which is equivalent to what we've done above) to work out full-scale sinking times. The sinking was quite complicated, and without experimental validation no conclusions could ever have been reliably determined. Theoretical results alone would not have been sufficient. Also, all model ship tests performed for working out full-scale wave drag use the same Froude number scaling process.

1.2 Scaling for operation *Chastise*

We've covered a sinking ship, so now let's look at something very different. During WWII, at attack was made on dams, and because this is a particularly instructive example of dimensionless analysis, so we'll consider it in a little detail. I've included this example because it has technical interest in terms of showing dimensional analysis used for an unusual problem.

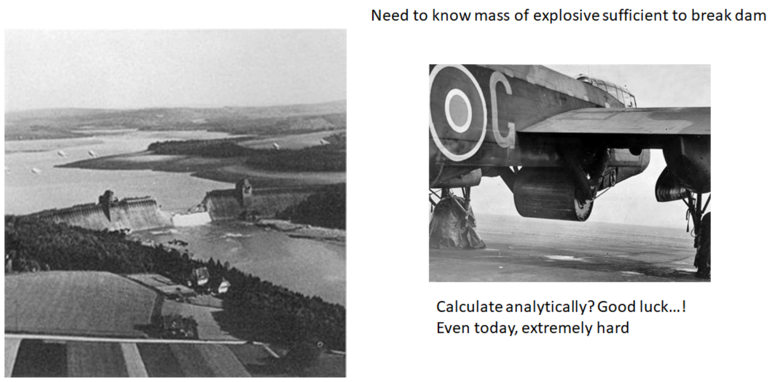


Figure 4: Operation *Chastise*

Figure 4 shows the overall arrangement, with the explosive charge designed to sink down the water face of the dam, while remaining in contact with it.

The dimensional parameters of interest are surmised as

- Energy E, kgm^{2}s^{-2}
- Maximum stress σ_{max}, Nm^{-2} of the dam structure
- length l, m (typically height of the dam)
- gravity g, ms^{-2}
- density of water ρ, kgm^{-3}

From this we can build

$$N_{destroy} = \frac{E}{\sigma_{max} l^3} = \frac{kgm^2 s^{-2}}{kgm^{-1} s^{-2} m^3} \tag{5}$$

$$N_{preload} = \frac{\rho g l}{\sigma_{max}} = \frac{kgm^{-3} m s^{-2} m}{kgm^{-1} s^{-2}} \tag{6}$$

Immediately, it is clear the energy required E, which is proportional to the mass of explosive, depends on the cube of the length scale. Achieving the same proportional damage on a dam twice the size will require eight times more energy. It is also clear from $N_{preload}$ than both a model and full-scale dam should be preloaded to the same fraction of their maximum stress before the test is conducted.

But what is the actual value of $N_{destroy}$? There is no easy way to know. To work it out analytically, you would have to solve the non-dimensional equations of motion for the dam and the water at the same time, during the explosion. Just knowing the dimensional scaling is not enough to know the

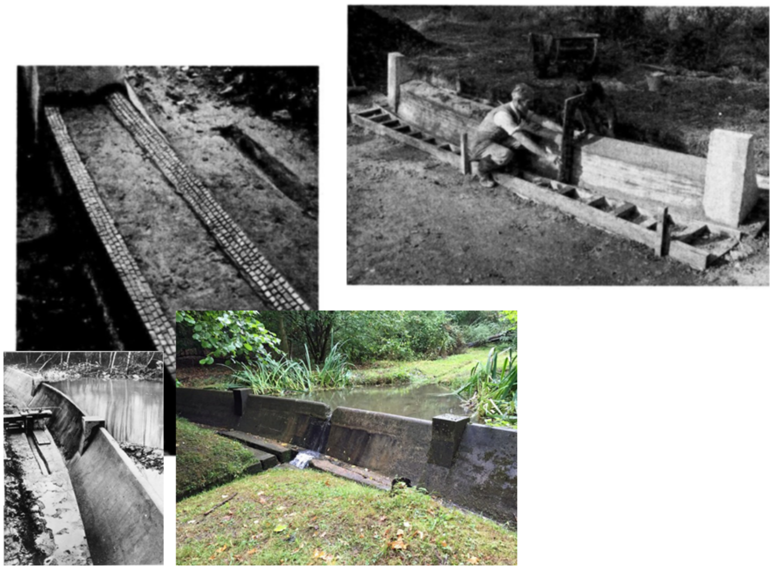


Figure 5: Small dam models

number itself. Those equations, however, almost certainly have no analytical solution, and making progress that way would be slow.

To resolve this, scaled experiments were deployed, as shown in figure 5. As you can see, these were quite tiny models. It's worth noting that these experiments were conducted by staff whose normal role was to assess blast damage on scaled buildings, so the scaling process was already well known to them. The models were made out of mortar, deemed to be similar enough to the actual dams (ie. similar σ_{max}), and early tests even used individual bricks in an effort to improve the representativeness of the results, although this was later deemed unnecessary.

The small tests indicated 28.3g of explosive was required at 50th scale, giving

$$28.3 \times 50^3 = 3544kg \tag{7}$$

at full scale.

These experiments, however, were so small that there was concern other non-dimensional numbers that had been ignored might still be important. So, a much larger disused dam in Wales at Nant-y-Gro was located, and corresponding models of this also built (including the shape of the lakebed behind it). These are shown in figures 6 and 7, with the final test in figure 8.

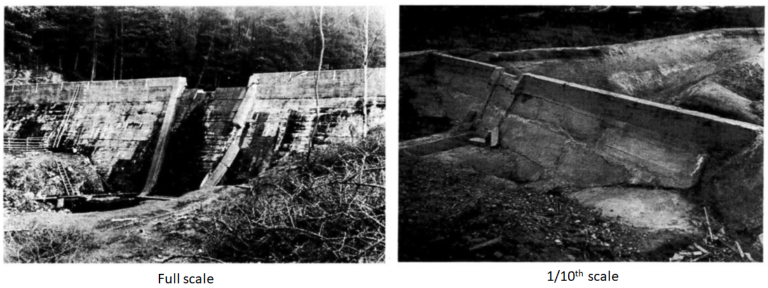


Figure 6: Full and scale models of the Nant-y-Gro dam

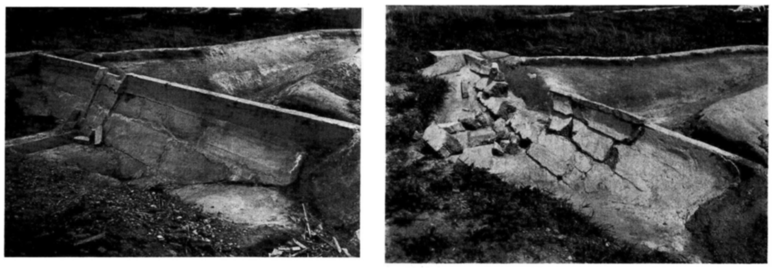


Figure 7: Model Nant-y-Gro dam before and after explosion

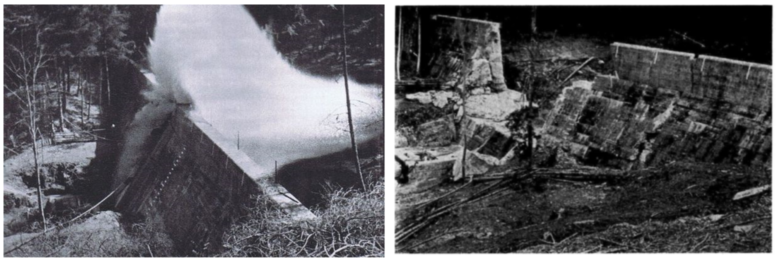


Figure 8: Full scale Nant-y-Gro dam during and after explosion

The model Nant-y-Gro dam was destroyed by 113g, which predicted 113kg would be needed for the full scale Nant-y-Gro. This was duly tested with 125kg (which happened to be available as a pre-existing charge), and the dam successfully breached. You can, in fact, go and visit the remains of this dam in the Elan Valley.

Nant-y-Gro was at 1:3 scale, roughly, of the real dams, which put the charge at

$$125 \times 3^3 = 3375kg \tag{8}$$

and was very close to the final 3000kg used. The results from this are shown in figure 9, and size of the hole compared to scaled predictions from the Nant-y-Gro test. The results are in reasonable agreement.

So, dimensional analysis is versatile and rapid, albeit at the expense of being somewhat tied to experiment. Perhaps most importantly of all, it takes away the question of scale. This is obviously useful for experiments, but it is also useful from a purely conceptual point of view in terms of design work, because it allows us to highlight the decisions that really matter.

1.3 Aircraft Icing

Let's also consider a third example, which is for aircraft icing. Parameters of interest can be surmised as

- liquid water content ρ_w, kgm^{-3}
- air density ρ, kgm^{-3}
- speed v ms^{-1}

Table 2. Forecast of dimensions of breach

Scale ratio*	Depth of charge: ft	Weight of explosive: lbs	Size of breach: ft	
			Depth	Width
1	10	280	24	60
2	20	2200	48	120
3	30	7500	72	180
4	40	18 000	96	240
5	50	35 000	120	300

* Related to Nant-y-Gro test

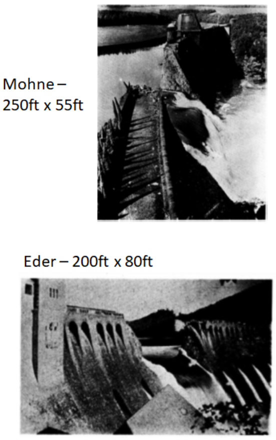


Figure 9: Final results

- time t s
- length l m
- mass of ice m_i kg

From this we can expect three non-dimensional numbers, the first being the non-dimensional mass of accumulated ice

$$p_{mass} = \frac{\rho_w l^3}{m_i} \tag{9}$$

and another which is the non-dimensional forward speed

$$p_{time} = \frac{vt}{l} \tag{10}$$

and also $p_{density} = \frac{\rho_w}{\rho}$ (which is almost invariably going to be the same at model and full scale, so it's not that important unless a different liquid were used).

Logically, we expect $p_{mass} = f(p_{time}, p_{density})$. Interestingly, the ratio of aerodynamic force to inertial force on water droplets is essentially p_{time}.

In principle, this type of approach allows icing experiments to be conducted a smaller values of v and l (as in a wind tunnel) to achieve representative results for full-scale icing. Also, ρ_w can be increased to provide a more rapid experiment. Once the function $p_{mass} = f(p_{time}, p_{density})$ is known any other combination of the dimensional inputs can be used to find the mass of ice accumulated in a given time.

The thermodynamics of ice formation were excluded, and if this were *included*, additional non-dimensional numbers would then appear.

That's it for non-dimensional numbers outside of aerodynamics! Hopefully this has given you some appreciation of how they can be used within engineering. There are, of course, many more examples to be found elsewhere.

Next, we're going to set dimensionless analysis in the context of aerodynamics.

2 Non-dimensional numbers in aerodynamics

Now let's looks at aerodynamic forces using the tool of dimensional analysis.

2.1 Force coefficients

The question of aerodynamic forces on aircraft of different sizes, flying at different speeds and air densities, and other physical dimensional conditions is completely answered through force coefficients. Of course, these numbers don't tell us the value of the coefficient; they just tell us how that answer changes with scale.

Starting with lift coefficient C_L (note subscript '3D' here to denote the three-dimensional definition, we will discuss 2D later)

$$C_{L_{3D}} = \frac{L}{\frac{1}{2}\rho V_\infty^2 S} = \frac{N}{Nm^{-2}m^2} = [0] \tag{11}$$

For drag

$$C_{D_{3D}} = \frac{D}{\frac{1}{2}\rho V_\infty^2 S} = \frac{N}{Nm^{-2}m^2} = [0] \tag{12}$$

It is worth remembering that $\frac{1}{2}\rho v^2$ has the dimensions of pressure, because it is $kgm^{-3}(ms^{-1})^2 = kgm^{-1}s^{-2}$, and pressure has dimensions of force per area $kgms^{-2}/m^2 = kgm^{-1}s^{-2}$. You can always remember the units of force, because it is a mass times an acceleration.

Note that using the same reference area for lift and drag is standard practice for aircraft, and means that

$$\frac{L}{D} = \frac{C_L}{C_D} \tag{13}$$

and to be frank we wouldn't want it any other way. You may wonder why we don't use 'frontal' or 'frontally projected area' and the main reason is that it's just not useful to do so for the aircraft as a whole; we already have a reference area (the wing planform area), so we use that.

This rule does change a little if we are looking to 'add on' component drag or lift values. For example, let's say you put a store, which is cylindrical with 'pointed' ends, underneath a wing. In this situation it is quite likely that C_D has been measured separately based on frontal area, or at the very least, measured relative to a reference area that was not your wing area (because the aircraft didn't exist when the store was built, let's say). In this case

$$D_{aircraft+store} = D_{aircraft} + D_{store} \tag{14}$$

$$C_{D_{aircraft+store}} \frac{1}{2}\rho V_\infty^2 S_{aircraft} = C_{D_{aircraft}} \frac{1}{2}\rho V_\infty^2 S_{aircraft} + C_{D_{store}} \frac{1}{2}\rho V_\infty^2 S_{store} \tag{15}$$

so that

$$C_{D_{aircraft+store}} = C_{D_{aircraft}} + C_{D_{store}} \frac{S_{store}}{S_{aircraft}} \tag{16}$$

and this is known as the 'drag buildup method', and you will use it in later years for your design project work. It's not really a method though; it's common sense. Note that in the real word drag doesn't just add up, because the drag of the total is greater than the sum of the parts due to interference drag.

Have you ever noticed that when you're driving, and a large vehicle overtakes you, you feel your car briefly decelerate, even though you didn't move the accelerator? This is intereference drag. The drag of two cars driving close together is greater than the sum of the drags with them far apart, because each car increases the flow speed around the other. The same thing happens with aircraft components, fuselages, wings, engines and so on.

That's for cars driving *side by side*. You can of course decrease your drag by driving close *behind* the car in front and sitting in its 'wake', which is effectively the air it is dragging along with it, but, erm, it's not a good idea at all.

We can also handle moments

$$C_{M_{3D}} = \frac{M}{\frac{1}{2}\rho V_\infty^2 Sc} = \frac{Nm}{Nm^{-2}m^2m} = [0] \tag{17}$$

where we need a further length c to cancel the dimensions. c is usually 'chord' or 'mean aerodynamic chord'. All that matters is that it is a length, but life is easier if it's a convenient one.

A common question is - could we get rid of the $\frac{1}{2}$ in the force coefficients, or could we use a different area S? Yes and yes. However, it would gain us nothing at all. The $\frac{1}{2}$ comes from integration of the momentum equations for a fluid (see later years) and although we could do without it here, the only effect would be to reduce the defined value of all coefficients by a factor of 2. We can use any reference area we want, from the area of the pilot's cap to the diameter of a can of tonic water in the duty free cart, **providing it scales with the aircraft**. Both of those reference areas scale with *people* rather than the *aircraft*, so they wouldn't be very helpful.

By now you're hopefully convinced that force coefficients are about both physics and convenience. It's absolutely shocking. Generation after generation of engineers using the same coefficients (with the exception of the originators of the idea, who thankfully delivered the whole system in the first place). Aren't we all a helpful bunch?

Think of engineers in the 22nd century not yet born - they will be using precisely the same force coefficients we use today, and will easily be able to interpret any data you generate during your lifetime by using the same force coefficient definitions applied here. The data you measure could be the data they use.

2.2 Force coefficients in 2D

Now, a perilous journey to a two-dimensional world. Sometimes we work with wings, but sometimes we just don't want to know about span and the third dimension. This brings us to *aerofoils*, and of course we have the same coefficients, but now we only have a length we can use rather than an area. However, forces L and D are now all defined in Newtons per metre (rather than Newtons), so we have

$$C_{L_{2D}} = \frac{L}{\frac{1}{2}\rho V_\infty^2 c} = \frac{Nm^{-1}}{Nm^{-2}m^2} = [0] \tag{18}$$

$$C_{L_{2D}} = \frac{D}{\frac{1}{2}\rho V_\infty^2 c} = \frac{Nm^{-1}}{Nm^{-2}m^2} = [0] \tag{19}$$

Not wanting to leave out the black sheep of the family

$$C_{M_{2D}} = \frac{M}{\frac{1}{2}\rho V_\infty^2 c^2} = \frac{Nmm^{-1}}{Nm^{-2}m^2} = [0] \tag{20}$$

where M is a moment per unit span, ie moment per metre span. However, a moment is a force times a length, so it just turns out that a moment per span is actually a force. If you see what I mean. Note that we have c^2 on the bottom now, because we needed an extra length to cancel dimensions again, and that's what the engineers of old did before us.

2.3 α, Reynolds number and Mach number

You've probably been wondering what exactly determines force coefficients. This is the neat part of non-dimensional system; *non-dimensional numbers can only depend on other non-dimensional numbers*. A change in the dimensional inputs can of course change the force coefficients, but only if they do so via changing one of the other non-dimensional numbers.

You're probably familiar with poking your hand (carefully) from a car window, twisting it, and feeling the lift force change. Clearly, then, angle ('angle of attack' or also 'incidence') is our first variable

$$\alpha = [0] \tag{21}$$

We know that for a circle circumference is $2\pi r$, where 2π is the angle in radians of a full turn. So, angles don't have units (because they are a ratio of two lengths). Hence, α is our first (and arguably most fundamental) non-dimensional input for finding force coefficients.

What comes next? There are two components. 'Rayleigh's method of dimensional analysis', which you may well have been taught, and 'Buckingham's π Theorem'. Rayleigh's method assembles all our physical variables raised to unknown powers, and then we solve simultaneous equations to find those powers. The π Theorem tells us how many non-dimensional numbers we are looking for, as $p = n - d$. These are in many ways the same thing - if we have n inputs and d equations (which cancel units in mass, length and time) then there must be only p truly independent variables (the non-dimensional parameters). Perhaps a more helpful way of thinking is to think about ratios of the relevant forces, because then we can remember what type of behaviour is represented.

I don't like either of these methods. Nothing is wrong with them, but I'm biased, and Rayleigh's method just gives you some numbers. It *can't* tell you which of the non-dimensional numbers matters the most, and it can't tell you if you've managed to include all the correct physical quantities in the first place. The π theorem is just an example in linear algebra and is more of an observation than a technique.

The Reynolds number represents the ratio between inertial forces and viscous forces, while the Mach number represents the ratio between inertial and the force required to compress an element of air. I'll now try to convince you of this.

There is a result you can't derive yet (wait until year 2 aerodynamics), which tells us that $\frac{dp}{d\rho} = a^2$ (where a is the speed of sound), so that $dp = a^2 d\rho$ and therefore compression force$= l^2 dp = l^2 a^2 d\rho$, or in dimensional terms just $l^2 a^2 \rho$. Inertial forces are mass times acceleration, hence $\rho l^3 \frac{V}{t}$, giving (remembering that $\frac{l}{t} \sim V$)

$$\frac{\rho l^3 \frac{V}{t}}{a^2 \rho l^2} = \frac{lV}{ta^2} \sim \frac{V^2}{a^2} = M^2 \tag{22}$$

which is the square of the Mach number. So, the Mach number tells us how large the forces of motion are compared to the forces required to compress the gas.

The Reynolds number can be found through a similar result, using the definition of viscosity $\tau_w = \mu\frac{\partial V}{\partial y}$, where τ_w is shear stress at the wall (having units of pressure) and y is distance away from the wall, giving

$$\frac{\rho l^3 \frac{V}{t}}{\mu \frac{V}{l} l^2} = \frac{\rho l \frac{V}{t}}{\mu \frac{V}{l}} = \frac{\rho l \frac{l}{t}}{\mu} \sim \frac{\rho l V}{\mu} = Re \tag{23}$$

Together with α, M and Re complete the triad of the most important non-dimensional numbers for determining aircraft force coefficients. α is the most important, because it always drives the problem, nomatter what Re and M are. M is the next most important; it makes a great deal of difference if it is above 0.3 (below that, you can ignore it). Re matters a lot if it is very small (below 50,000), but causes only fairly gradual changes above 10^6. In between those two levels, it causes a variable level of mayhem (boundary layer transition and changes in separation behaviour), but effects can be mitigated in wind tunnel tests.

A good example of the type of relationship betwen lift coefficient and angle of attack is given in figure 10, for the 747-100/200. You can see how lift increases, before the curve flattens out. Different lines are shown for different flap angles. Notice that there is not always a drop in lift, but if you looked at the drag curve you would see a very sudden increase there as separation began. A more obvious drop in lift is seen with the higher flap angles. $C_{L_{mac}}$ looks to be about 2.5 for the flaps 30 degrees case (although flap 'angle' is a somewhat meaningless variable for the triple slotted system on the 747).

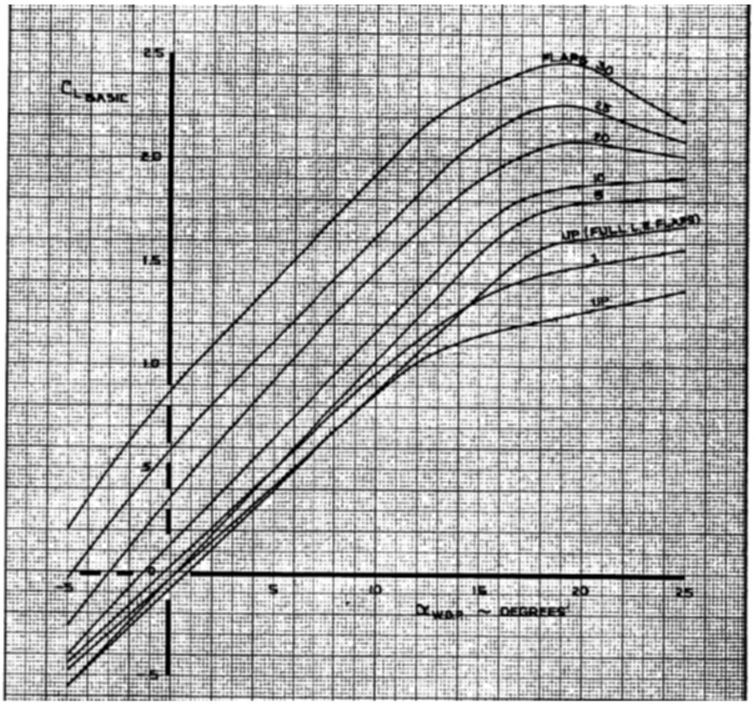


Figure 10: 747 C_L vs. α curve

2.4 Wind Tunnel Testing

So what can we now achieve, armed with a knowledge of force coefficients, α, Re and M? Answer - *everything*.

Owing to the fact that the system we are using now has no dimensions, you can argue that it doesn't matter what size the aircraft is, so it can just as easily be a wind tunnel model as the full scale aircraft. Life isn't quite that easy, though, and there are a few more things to think about:

1. The force coefficients are only correct for the full scale aircraft if the angle of attack is the same for the model and the full scale aircraft - wow, well that's easy, just set them at the same angle
2. The force coefficients are only correct for the full scale aircraft if the Mach number is the same for the model and the full scale aircraft - not so easy, but in principle we can run the tunnel at the flight Mach number (which might be quite high)
3. The force coefficients are only correct for the full scale aircraft if Reynolds number is the same for the model and the full scale aircraft - uh-oh, not easy at all!

This all boils down to the idea that the non-dimensional outputs only match at different scales if the non-dimensional inputs are the same. If the real aircraft is flying at $\alpha = 5^o$ then the model needs to do the same, and M and Re need to match, otherwise the non-dimensional outputs (C_L, C_M, C_D) will be invalid.

The Mach number can always be matched by running the tunnel at the flight Mach number. Unfortunately, $Re = \frac{\rho V l}{\mu}$, so if we increase V then Re changes. Worse than that, if l is reduced (such as when building a 1/10th scale model), then Re will be far too small and increasing V to compensate will leave M 10 times too big. The most common approach in tunnel testing is then not to try and force Re to the correct value, which is impossible in most practical cases, but to adjust the model with devices that mimic behaviour that would be seen at higher Re, such as trip strips for triggering boundary layer transition to a turbulent state.

The exception to this is if you have a big enough budget. There are currently two well-known tunnels in the world which can match flight values of M and Re for large aircraft, these being the National Transonic Facility (NASA) and the European Transonic Windtunnel (ETW). Both these are pressurised cryogenic tunnels, which achieve flight Re by increasing pressure (to raise ρ) and dropping temperature (to lower μ). They are exceedingly expensive to operate, and models are also costly because they must withstand large forces and extreme temperatures. Other more secretive pressurised cryogenic tunnels probably also exist.

As a footnote, it is worth pointing out that transonic tunnels were the hardest to build successfully. It was not until post-WW2 that the importance of using ventilated sections was widely appreciated; ventilated test sections allow the static pressure in the tunnel to be varied near the model in a controlled manner, giving very accurate control of M there. If this is not done, tunnel wall effects make it hard to adjust M reliably, because flows near $M = 1$ tend to be very sensitive to the combined shape of the tunnel and model.

2.5 Aeroelasticity

A slightly more complicated example of how further non-dimensional numbers become involved is through aeroelasticity. The ratio of dynamic pressure to Young's modulus, given as $\frac{\rho V^2}{E}$, measures how flexible a structure is when exposed to aerodynamic forces, and is often known as 'speed index'. This ratio changes all the force coefficients, as you can judge by figure 15. Clearly, the loads on the wing changed the shape substantially, which in turn alters the loads again. This interaction between aerodynamics and elasticity is known as *aeroelasticity*. It can exist in both static (non-moving) and dynamic (moving) cases. The figure shows the static (ie. non-moving) deflection of a 747 wing; 150+ tons bends it quite a lot!

(a) Undeflected (b) Deflected

Figure 11: 747-400 wing before and during flight

3 Stability and Trim

Stability and trim seem to manage to be two of the most commonly misunderstood topics relating to aircraft, second only to the origin of lift itself. A non-technical interpretation of 'stability' is often based on our human experience of balancing while standing up, riding a bicycle or getting in to a boat. Although related, most of this is misleading or irrelevant in terms of aircraft.

'Stable' refers to a system which when nudged or 'perturbed' from its initial state will move so as to return to that original state. The return to the initial state may follow an oscillatory or non-oscillatory path, but the system will nonetheless head back there. Similarly, an unstable response may or may not oscillate, but it is characterised by departure from that initial condition.

Aircraft have both static (non-moving) and dynamic (moving) stability about all axes.

I was just proofreading and realised how little sense that comment makes - isn't all stability about moving things?

The answer is yes - but also no. Static stability is when something is stable or unstable even before it has begun to move. Dynamic stability is when a *motion itself* is either growing or decaying. A stationary bicycle is statically unstable, but a moving bicycle is dynamically stable proving the rider's inputs are considered part of the system. Of course, it would be dynamically unstable without the rider's inputs, which is just a manifestation of the underlying static instability. You see what I mean? Perhaps stability is confusing afterall.

However, the most elementary and critical aspect of stability for aircraft is in pitch (nose up and nose down). This doesn't mean stability in other axes is unimportant, but just that pitch 'longitudinal' stability is a prerequisite. If that condition is not met, it will be largely irrelevant what happens in terms of the others.

We're also going to restrict ourselves to *static* stability. Longitudinal dynamic stability is also really interesting, but beyond AVDASI 1's scope. Most aircraft, with some exceptions, must be statically stable. Dynamic stability can often be helped via control inputs from the pilot.

Perhaps the most obvious aerodynamic stability people see relates to weathervanes. The lateral lift force behind the pivot of the vane always orientates the nose to point in to wind (see figure 12). Aircraft too have similar behaviour in terms of the forces acting on their tailfin, but there is no pivot, with the centre of mass becoming the relevant reference point because this is the location about which accelerations generate no moments. We could use another reference point apart from the CG, but then we would have to include inertial moments. Weirdly, the world's largest weathervane (so-I'm-told) is an aeroplane, a DC-3, mounted on a pole at the Yukon Transportation Museum.

Perhaps it's no surprise then that aircraft flying crosswind takeoffs and landings will tend to 'weathervane' in to wind all the time, which is compensated for using the rudder and undercarriage steering, and the art of accomplishing this makes for many entertaining youtube videos during the UK's winter storms. All these aircraft have strong stability in yaw, but what is going on is that the sideways wind is altering the equilibrium yaw angle, or yaw 'trim'. The trim point is the one where there is no resulting yaw moment - ie. the aircraft is at a steady, fixed yaw angle (there is no yaw rate). The pilot's goal is to ensure this yaw trim point is acceptable relative to the runway direction.

However, the most important behaviour by far, because it makes the difference between success and failure in terms of flight, is the stability in pitch. Quite obviously, if an aircraft has a natural tendency

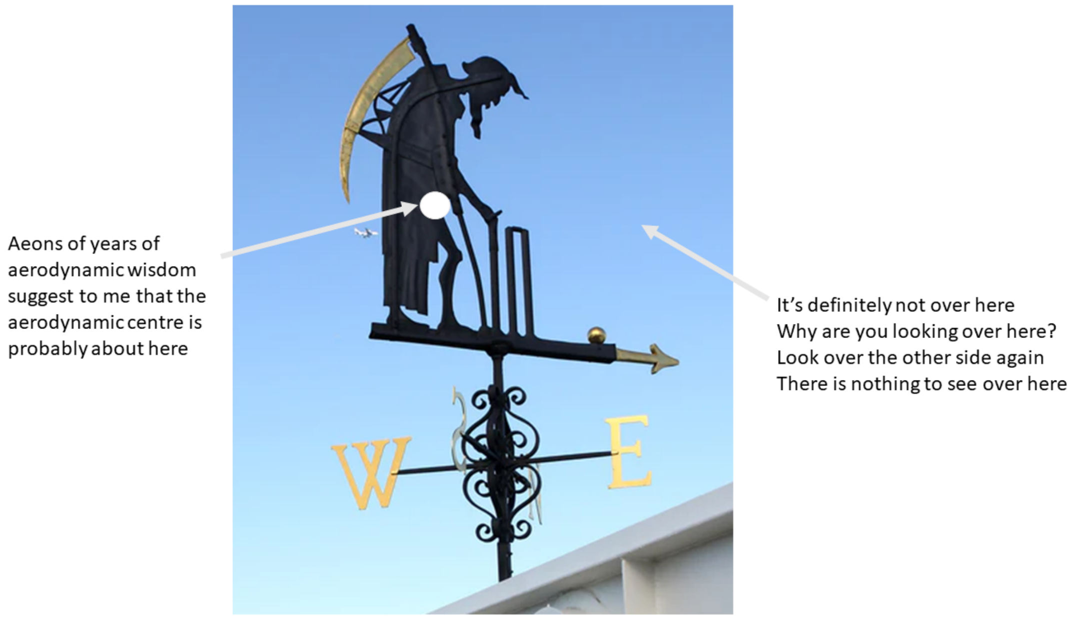


Figure 12: This weathervane is meant to be Old Father Time, but could equally well be a weary course lecturer. Note the aircraft in the background!

to flip nose up or nose down, without any control input, this will be a difficult or impossible aircraft to fly. One can consider the weathervane again in this regard - if the pivot of the vane is moved backwards, it will move more slowly in to wind. Move the pivot far enough back, and the vane will not point upwind, and move it further and it will swing right round and point downwind. For an aircraft, which is not pivoted (except at that Canadian museum I mentioned), the CG replaces the pivot in principle.

Behaviour in pitch is therefore dominated by three properties: (i) the CG location, (ii) the location where the change in wing lift force acts for small changes in angle (wing aerodynamic centre) and (iii) the location where the change in tailplane lift force acts for small changes in angle (tailplane aerodynamic centre). We can actually group the wing and tailplane lift forces to a single location, called the 'neutral point' instead, to help simplify things. The only thing that then matters is the distance from the CG to the neutral point, a distance which is normally known as the 'static margin' and expressed as a % of mean chord.

However, before considering 3D lifting surfaces, let's first look in 2D. So, where does the lift force on an aerofoil act? This brings us to the thorny issue of centre of pressure and aerodynamic centre, two locations which are very different.

- The aerodynamic centre x_{ac} is the location about which $\frac{\partial C_m}{\partial \alpha}$=0
- The centre of pressure x_{cp} is the location about which $C_m = 0$

In turns out that x_{ac} remains very close to 25% of chord for nearly all aerofoils in 2D. However, because aerofoils can generate moments even when the lift is zero, x_{cp} moves around a lot.

The 25% result not new. It was only formally derived in the early 20th century as part of *thin aerofoil theory* which you will meet in year 2, but has been known for centuries (by my guess). Windmills often used a spar position near 30%, and noting that the 3D postion is often behind 25% for lowish aspect ratios, this makes perfect sense, as in figure 13. The spar positioned here means there will be little twisting under loads. Similar principles apply to ship rudder design, where these were often balanced with a pivot near 25% to make them easier to move. Perhaps the real question is why it took the theory so long to catch up with what people already knew?

Figure 13: A windmill with interesting spar positioning

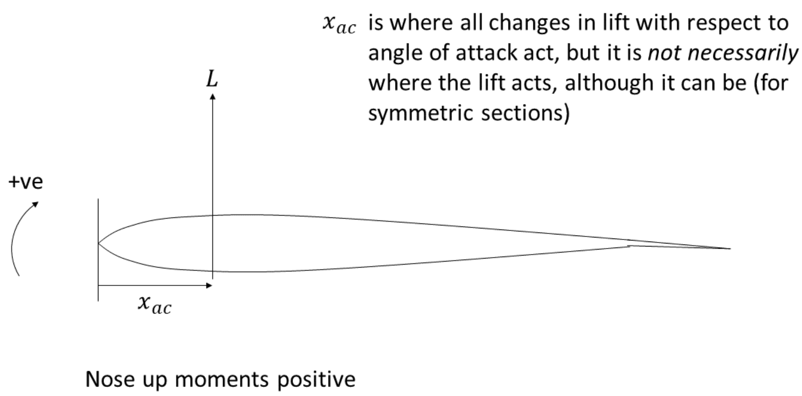


Figure 14: A humble aerofoil

Now consider (with C_m defined about the leading edge, and nose-down negative) (see figure 14)

$$C_m \frac{1}{2}\rho v^2 c^2 = -C_l \frac{1}{2}\rho v^2 c x_{ac} + C_{m_0} \frac{1}{2}\rho v^2 c^2 = -C_l \frac{1}{2}\rho v^2 c x_{cp} \tag{24}$$

where C_{m_0} is the moment that would exist anyway if the lift were zero. Of course, standard practice is to cancel through by dynamic pressure and a length or length squared, ie. $\frac{1}{2}\rho v^2 c^2$, so

$$C_m = -C_l \frac{x_{ac}}{c} + C_{m_0} = -C_l \frac{x_{cp}}{c} \tag{25}$$

which is a non-dimensional equation - neat! Now clearly

$$\frac{x_{cp}}{c} = -\frac{C_m}{C_l} \tag{26}$$

If $C_l = 0$, which does not imply $C_m = 0$ (but only that $C_m = C_{m_0}$), then $\frac{x_{cp}}{c}$ is undefined - in effect, it is infinitely far from the aerofoil, trying to find an infinite moment arm to produce a moment with zero force. This makes it an unhelpful reference location. It is also clear that $x_{ac} = x_{cp}$ if $C_{m_0} = 0$, which is true for any symmetric aerofoil. The fact that it is possible to have $x_{ac} = x_{cp}$ is one of the things that leads to confusion. We like things to always be the same or always different, but when they're often different but peskily sometimes also the same, it upsets our inner sense of truth.

Now differentiate to get

$$\frac{\partial C_m}{\partial \alpha} = -\frac{\partial C_l}{\partial \alpha}\frac{x_{ac}}{c} \tag{27}$$

That constant C_{m_0} has gone courtesy of differentiation, leaving a neat definition of x_{ac}

$$\frac{x_{ac}}{c} = -\frac{\frac{\partial C_m}{\partial \alpha}}{\frac{\partial C_l}{\partial \alpha}} \tag{28}$$

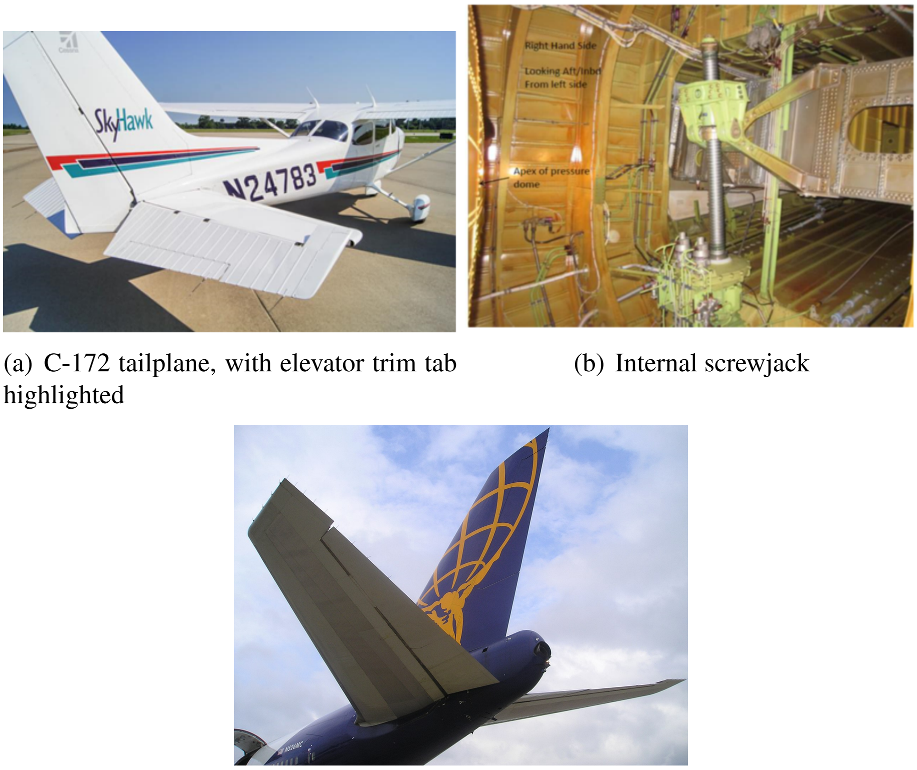
So clearly x_{cp} is given by the moment/lift ratio, and x_{ac} is given by the rate of moment/rate of lift ratio with respect to angle of attack.

It isn't quite true to say that x_{ac} is always fixed. The differentiated expression is normally considered about angles of attack where $C_m = 0$, ie. trim points (not the same as $C_{m_0} = 0$...). Trim points at low angles of attack show predictable behaviour, but for a trim point involving separated flow then x_{ac} would move to a different location. Generally, though, you won't encounter analysis at these sorts of conditions during your degree.

An even more obvious movement in x_{ac} takes place with changes in Mach number. 25% is only true for speeds below transonic ($M < 0.7$ roughly), after which x_{ac} moves around quite a lot due to appearance and movement of shocks, before eventually settling on a new home at 50% chord for speeds above transonic ($M > 1.4$ roughly). Actually, this is a really important point, and unfortunately quite a few test aircrew died in the early post-WW2 period discovering this fascinating feature of the flow equations. As M increased, nose-down moments exceeded the maximum nose-up moment of the pitch control, with dire consequences. This was rectified with larger, all-moving tailplanes, improved tunnel testing, and engineers generally getting their act together (perhaps that's not fair - they were under a lot of time pressure to build faster aircraft, afterall).

3.1 3D values of x_{ac}

The foregoing arguments were all for 2D aerofoils. Unfortunately there are no particularly simple arguments for the location of x_{ac} in general for 3D wings, because chordwise loading changes depending on spanwise position. However, a location for x_{ac} still exists, although it's not 25% any longer unless the aspect ratio is very high and the planform rectangular. The lower the A_R, the further aft x_{ac} is. If you think about very long, thin shapes like fuselages or javelins (with $A_R << 1$), logic dictates x_{ac} will end near 50% chord again, which is borne out in practice.



(a) C-172 tailplane, with elevator trim tab highlighted

(b) Internal screwjack

(c) Tailplane (the horizontal unpainted surface)

Figure 15: Trimtabs, 747 screwjack mechanism and 747 tailplane

A position for x_{ac} still exists for a 3D surface and is critical in terms of stability. Calculating exactly where it is requires some more elaborate methods you'll meet in year 2, or use of experimental data, but is not fundamentally complex. Everything else, however, remains unchanged.

When I was building a small delta winged glider recently, I built a $\frac{1}{2}$ scale test model to find x_{ac} through trial and error - which is almost embarrassing to admit! However, it was quick, cheap and accurate to do it that way. For a large expensive aircraft, you would use a computational method, then check it against tunnel tests, and finally confirm it later from flight test measurements.

3.2 Grand longitudinal stability conclusion

NB - this section is from the 'glider task' pdf, which I do suggest following if you'd like a more practical understanding.

With these ideas, we can find the aerodynamic centre (or 'neutral point') of a wing+tailplane configuration as in figure 17. Let's linearise each force as being a gradient multiplied by the derivative with respect to α, and consider constant moments as a separate contribution (which cancels). Also $A_{ref} = A_w + A_t$.

Taking moments of aerodynamic force about a reference point for the wing and tailplane, and equating this to the moment of a total lift force acting at $x_{ac_{w+t}}$

$$(C_{L_{\alpha_w}} A_w + C_{L_{\alpha_t}} A_t)\alpha x_{ac_{w+t}} + C_{m_0} A_{ref} c_{ref} = C_{L_{\alpha_w}} \alpha A_w x_{ac_w} + C_{L_{\alpha_t}} \alpha A_t x_{ac_t} + C_{m_0} A_{ref} c_{ref} \quad (29)$$

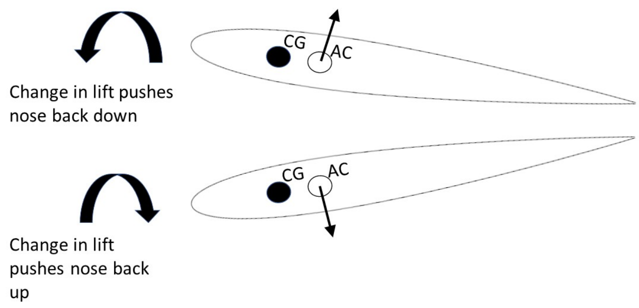


Figure 16: The basic premise of longitudinal static stability

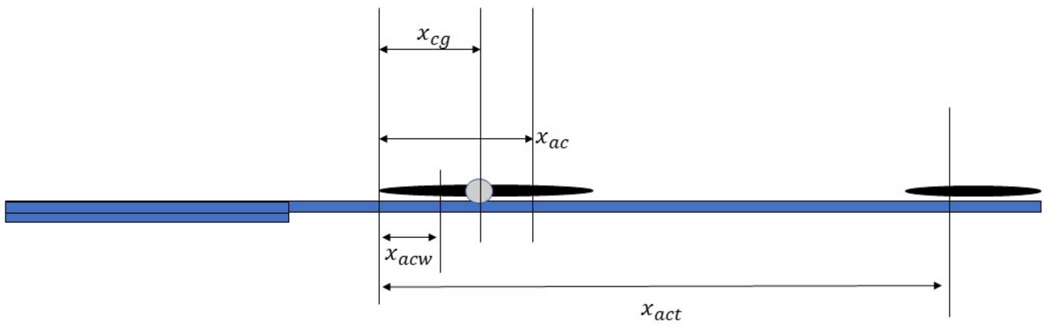


Figure 17: Reference locations on wing and tailplane

cancelling the C_{m_0} terms and assuming $C_{L_{\alpha w}} = C_{L_{\alpha t}}$ (ie. lift gradients of wing and tailplane are the same) then x_{ac_t} is simply an area weighted addition

$$x_{ac_{w+t}} = \frac{A_w x_{ac_w} + A_t x_{ac_t}}{A_w + A_t} \tag{30}$$

Note that a constant moment C_{m_0} was included to show that it isn't important in determining the aerodynamic centre location for the overall configuration. Also, dynamic pressure was cancelled across all terms.

There are some important restrictions that means this convenient and instructive result needs to be used with a little caution. These are:

- The assumption that $C_{L_{\alpha w}} = C_{L_{\alpha t}}$ is not true if the aspect ratios of wing and tailplane are quite different, and so I'd ask you don't use the result above to design anything other than model aeroplanes
- The angle of attack seen by the tailplane is actually influenced by the downwash (downwards flow) behind the wing. More accurate estimates will include this effect - please refer to your flight dynamics notes in later years

Lift gradients depend on the aspect ratio of the surface, and they will only be the same if both surfaces have high aspect ratios. In this case the tailplane will likely show a lower gradient than the wing (for the 10g glider, an AR of 6 for wing versus 2.6), although the twin fins on the tailplane of the 10g glider will mitigate this, acting as end plates/winglets. If we don't make this assumption we will end up in a rabbit hole worrying about lift gradients, so we shall make it for now, and you can come back to it in aerdynamics 2. It means our $x_{ac_{w+t}}$ locations will be bit too far backwards.

Equation 30 shows that (roughly) the neutral point is located at the area-weighted average position of the wing and tailplane aerodynamic centres. This makes sense, because we've assumed they have the same lift gradients. Keep in mind this is also strictly for low M; at high M the gradients change in a complicated manner for the wing and tailplane separately, moving $x_{ac_{w+t}}$ according to a different relationship. So, this expression is useful for teaching, but for real design you would want to refine it. However, based on some tests moving CG around, it's actually quite accurate for the 10g glider.

Most importantly, if the CG is located at the neutral point, then there is no rate of change of C_m with respect to α (neutrally stable). If the CG is in front of this point, a nose up change in α gives a nose down change in moment (statically stable). If the CG is behind this point, a nose up α gives a nose up moment, which gives static instability.

3.3 Canards and tail-less aircraft

Canards are aircraft with a *foreplane* rather than a tailplane. An early canard aircraft, the 14-bis, resembled a duck, and the name stuck, canard being French for duck.

There is a tendency to believe that canards, shown in figure 18, are unherently unstable. However, this belief is *completely incorrect*. There is not an ounce of truth in it. Using a foreplane instead of a tailplane does indeed move the neutral point of the aircraft further forward than if a tailplane had been used, and you can see this would be true from the area weight calculation for neutral point

(a) 14-bis - supposedly the one that looked like a duck

(b) Saab Viggen - I don't personally think it looks like a duck, but perhaps you've met stranger ducks than me!

Figure 18: Some canards

location. However, all that subsequently matters is then the CG location relative to the new neutral point location. The principle of CG location for stability is unchanged; a canard can be stable or unstable just like a conventional design depending on where the CG is located.

A further question arises regarding aircraft that have no tailplane, such as Concorde. Again, there is no mystery here. The aerodynamic centre of the wing is the neutral point, because there is no tailplane, and stability depends only on CG location relative to this point.

The tail-less aircraft shows quite neatly the difference between centre of pressure and aerodynamic centre. The aerodynamic centre remains behind the CG for stability, but in equilibrium, the centre of pressure sits on top of the CG, otherwise the aircraft would not be in equilibrium. The controls on the trailing edge (elevator/ailerons, known as elevons) move the centre of pressure around, generating pitch or roll moments, but they are incapable of moving the aerodynamic centre, which remains fixed at a position behind the CG.

So neither canards nor tail-less aircraft pose any challenge to the principles of trim and stability we have discussed. You can build them out of paper and card, as we have often done as children, and they fly perfectly well, just like any other aircraft. Only minor point - the elevators on a canard go down to pitch the aircraft up - but that's a small difference! There are further details to canards and tail-less designs that make them preferable in some cases, but they are unrelated to stability.

If you still don't believe me, just build some! The best way to learn how aeroplanes fly is to make some that do, and more that don't.

3.3.1 Supersonic aerodynamic centre

Something that probably didn't occur to the designers of early windmills was what would happen at supersonic speed. Subsonic aerodynamics is governed by an elliptic differential system, sonic by a parabolic system and supersonic by a hyperbolic system. What this means is that at low speed acoustic waves can travel everywhere, at the speed of sound they exactly cannot move upstream, and beyond the speed of sound they are convected (blown) backwards within a particular cone.

In a practical sense, this alters how much of the geometry is 'seen' by the air, which alters the flow radically. Many things change, but importantly the aerodynamic centre for a 2D section moves from 25% of chord back to 50%. The change is in fact gradual and complicated, and takes place from about

Figure 19: F-16 all moving tailplane - ‘tailerons’

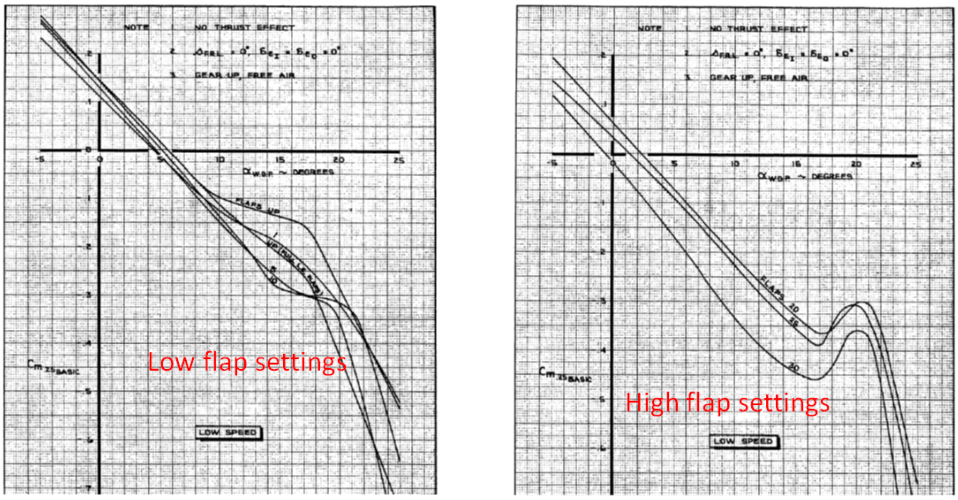


Figure 20: Variation of C_m with flap setting for 747

Mach 0.7 to 1.4. It means that any aircraft designed for subsonic flight will become substantially more stable at supersonic speed. There is also a substantial nose down moment, because the general position of the centre of pressure also moves aft.

99% of aircraft, if supersonic at all, are only designed for short periods of supersonic operation, especially military aircraft. Even military aircraft designed with ‘supercruise’ will not be supersonic for more than an hour (my guess). Therefore, they deal with the motion of the aerodynamic centre through (i) planform design and (ii) have plenty of trim authority (eg. all moving tailplanes). All moving tailplanes as in figure 19 are used almost universally. An example of the large changes in C_m what occur with flap angle on a commercial aircraft like the 747 is given in figure 20.

Only Concorde and the Tu-144 were built for sustained supersonic flight. It was not long after entry in to service that Concorde surpassed the total hours of all supersonic flight ever made up to that point in history. Sadly the Tu-144 design adopted somewhat unsuitable engines, was beset with system faults, and never fully achieved performance goals with fairly limited supersonic hours achieved.

To achieve reasonable supersonic efficiency, Concorde used pumping of fuel to move the CG fore and aft. The CG moved aft when supersonic, to avoid excessive nose down moments and high trim drag, and was moved forwards again when decelerating. Rapid fuel pumping was possible to move the CG forwards if an emergency descent was needed. Failure to move the CG forwards again would leave the aircraft with insufficient stability when subsonic, so this was a critical system. In addition, the delta planform was able to reduce the aero centre motion somewhat.

The fuel pumping process is shown in figure 21. This is almost a unique system, and would be an important part of any supersonic delta design. Note that variable geometry designs are able to move the aerodynamic centre, so the effects are not as pronounced on those aircraft (eg. F-14, Tornado, F-111). Variable geometry for an aircraft of Concorde's size would have been too heavy and reduced payload.

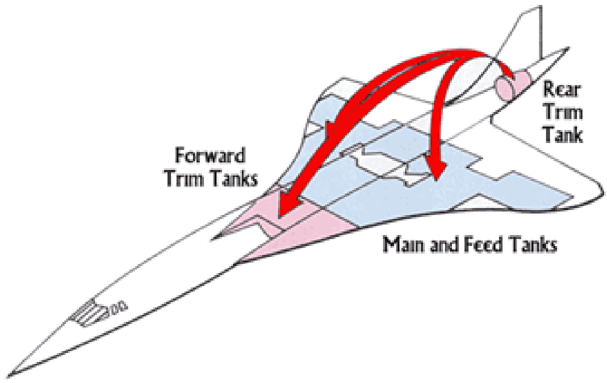


Figure 21: Concorde fuel pumping

4 Control and controls

4.1 Controls for manoeuvring

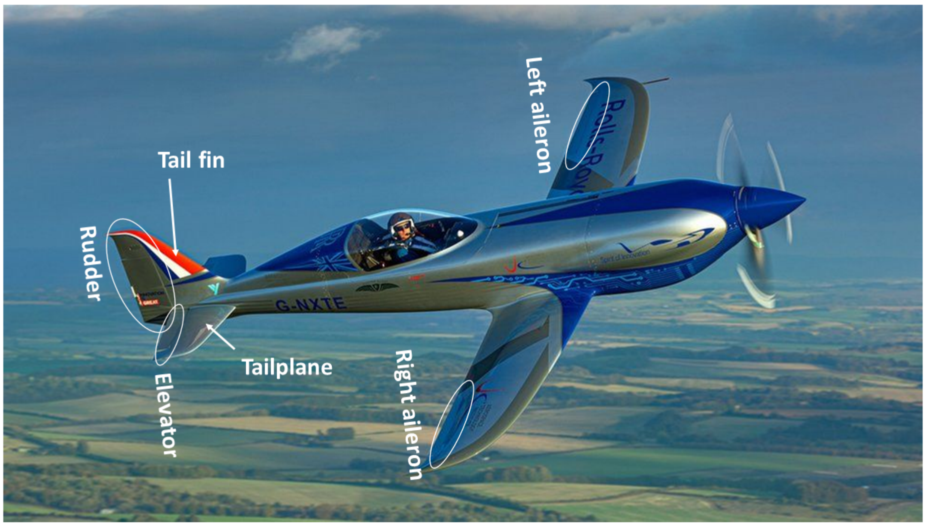


Figure 22: Some very standard controlled surfaces labelled on Rolls-Royce's attractive *Spirit of Innovation* electric aircraft, which at time of writing was the world's fastest electric aircraft (330mph). This one definitely looks suitable for an engineering academic commuting a few miles to the office

As illustrated in figure 22 aircraft require control about the three primary axes of pitch (controlled by the elevator), roll (controlled by a pair of ailerons) and yaw (controlled by the rudder). These are the typical controls, but others such as spoilers or tailerons can also be used, so not all aircraft are identical in this regard.

Of these, pitch is the most important (being the choice between ground and sky), followed by roll and then yaw. Most aircraft can be controlled/turned using pitch and roll only (large wingspan aircraft like gliders will require yaw coordination to avoid messy adverse yaw), and those with strong dihedral effects (mostly entry level RC planes) can be flown with pitch and yaw, because when there is a strong dihedral effect yaw produces roll quite quickly.

So, it may be surprising to learn that yaw, usually implemented with the rudder, is a somewhat secondary control. It is certainly important in many specific cases (eg. engine-out for multi-engine aircraft, directional control on the ground, crosswind landings, stall turns) but it is not crucial for much manoeuvring. With the obvious exception of gliders, most light aircraft can be flown through turns without rudder coordination, and only the lightest touch of rudder is needed to counteract adverse yaw from the ailerons. Gliders are an exception because the induced drag on the upgoing wing, caused be deflection of the aileron downwards to increase lift and produce roll, has a large moment arm about the CG. This yaws the glider outwards from the turn, is uncomfortable and increases drag further. For these reasons coordination is preferred, whereby a rudder input prevents the adverse yaw. Go and fly a glider and see for yourself!

But I digress - pitch, roll, yaw. Aircraft turn by rolling until the lift vector points towards the centre of the turn, and then the aircraft is pitched to increase the lift to provide a centripetal force component. Aircraft do not turn by generating centripetal force on the tailfin/rudder - it is far too small to achieve that, and also positioned in the wrong place for that to work well. It's the wings that turn aircraft.

Any lifting surface can change its lift coefficient by either (i) adjusting a flap on the trailing edge (ii) altering its angle of attack or (iii) dumping its lift by deploying a spoiler on the top surface, or increasing it with a spoiler on the lower surface.

- The trailing edge plain flap type control is very common. It's effective - remember, the control changes the pressure distribution on the whole of the aerofoil, not just the flap part, so it generates much larger forces than you might expect. At higher speeds, the forces on the control can cause the wing to twist the 'wrong' way, which means it's less desirable at higher speeds unless sufficient stiffness exists - see later discussion of inboard/outboard ailerons
- All-moving controls are usually found on the tailplanes of military aircraft. The desire for high speed, high roll and pitch rates and transonic flight with large moment changes tends to drive this. These systems are somewhat heavier, as all loads must pass through a pivot - the only exception being the wing-warping approach of the Wright brothers. However, this is not currently used on large aircraft, as it would require low stiffness, which would introduce other problems in terms of aeroelastics
- Spoilers are usually associated with commercial aircraft landing, where they open to improve braking effectiveness by dumping lift quickly once on the ground. However, the humble spoiler *also* provides roll control in flight through differential use, and has the benefit that it adds drag on the downgoing wing, which tends to balance adverse yaw from ailerons. Some aircraft, eg. F-14, Tornado and other fast jets, only have spoilers on their wings for roll control, and no ailerons. At high speeds there is an all-moving tailplane, which can move differentially to provide a roll input. I'm only aware of one use of spoilers for pitch control, which was on the V-1. The final dive was initiated with a spoiler on the bottom of the tailplane, which increased lift, gave a nose down moment and produced a dive

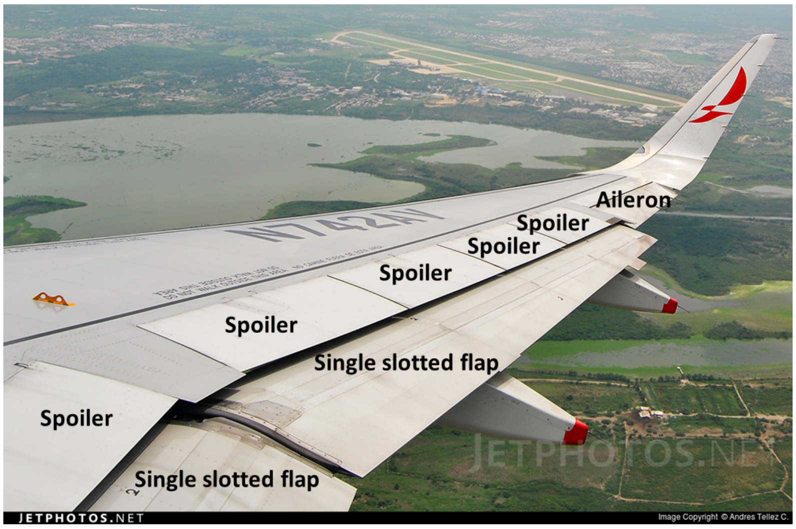


Figure 23: An A320 wing

4.2 Examples of other controls

There's a slight danger of thinking all aircraft use straightforward elevator-aileron-rudder controls. Of course, this is untrue, and some interesting alternatives are in figures 24 and 25.

The example for the L1011 is one of ‘direct lift control’. Using this technique, on approach to land spoilers were slightly extended on the wings, and the aircraft trimmed to fly the approach in this condition. Then, to control the aircraft above and below the flight path or ‘glideslope’ to the runway, the spoilers were extended a little to descend more or retracted a little to go up.

You may wonder why this was done, when the same could have been achieved by controlling the elevators. The answer lies in ‘pitch transients’. When the elevators deflect up, the first thing that happens is the force on the tailplane increases in the downwards direction, which means the aircraft *actually descends* before going up again. This transient is accounted for by pilots at a human level, but for the autopilot that Lockheed built for L1011, using the spoilers in this way permitted greater height accuracy for approach and autoland. The aircraft could be flown manually using this type of control if the pilots wished, or it could be disabled for manual flying if desired.

Another example is on the 777 rudder (although this type also exists on the DC-10, 727 and others). This is a ‘two-stage’ rudder, which is effectively a rudder with a flap on the trailing edge. The purpose is to increase rudder sensitivity, and also to increase the maximum lift on the tailfin for engine-out cases. The second part of the rudder is mechanically geared, so it doesn’t require any additional actuators, and this is a useful trick to improve performance of the tailfin without increasing its area (and thereby increasing weight and drag).

The F-14 is also pictured without any ailerons at all. Roll control is via spoiler and ‘tailerons’, whereby the all-moving tailplane sections move in different directions (a bit like what you would do to make an arrow spin). This works well at high speeds when the wings are swept, and at low speeds with the wings forward the spoilers are effective instead.

So, my point here is that aircraft controls are actually rather more interesting that the aileron-elevator-rudder ‘standard’ setup. The details of control are driven by the design objectives of the aircraft in question, and many configurations are possible. Perhaps you can think of some others?

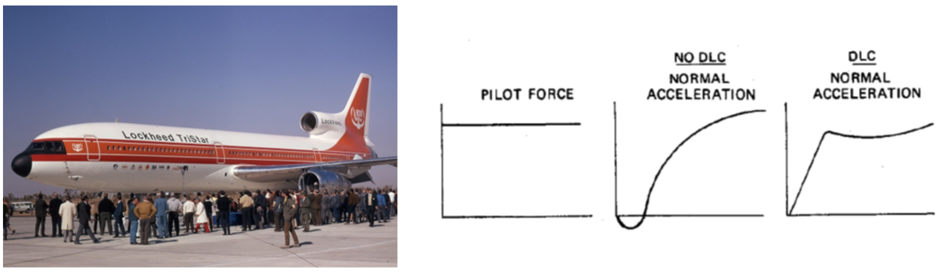


Figure 24: The L1011 and direct lift control influence. ‘Normal’ acceleration refers to up/down acceleration

4.3 High-lift devices

Stall speed of an aircraft from an equilibrium argument is $\sqrt{\frac{2\frac{W}{S}}{\rho C_{L_{max}}}}$, so apart from lowering the wing loading (leaving cargo behind or increasing wing size), or seeking climates where ρ is larger (low altitudes, cold days) the only route to lowering stall speed is to raise $C_{L_{max}}$. Common $C_{L_{max}}$ values are around 1.5 for clean wings (or below 1 if Re is low) rising to about 2.5 for a flapped wing with leading edge devices.

(a) F-14 - notice the absence of ailerons

(b) Two-part 'flapped' rudder on 777 tail-fin

Figure 25: Some varying configurations

The influence of high-lift devices is shown in figure 26. There is a limit to what can be achieved; $C_{L_{max}}$ above 3 is unlikely unless something dramatic is done, such as using blown flaps (eg. F-104 or Buccaneer).

The reason why a TE flap raises the lift curve at fixed α is straightforward. A flap essentially increases camber, and produces a greater momentum exchange with the air, leading to a larger lift force. What is not at all simple is why there is a change in $C_{L_{max}}$, and this is a value also of high importance.

The ultimate reason for this is that a flap produces a larger C_L value for the same adverse pressure gradient behind the suction peak. This adverse pressure gradient is an important factor in fixing when stall/separation occurs, with stall essentially occurring at a particular magnitude of adverse pressure gradient. Reducing the adverse PG for a particular C_L therefore leads to an improvement in $C_{L_{max}}$. For more information, see the A1 lab in aero 2 in year 2. Similar arguments can be used to justify the influence of a slat or droop on $C_{L_{max}}$.

It's worth pointing out that of course these devices in many cases also increase the wing area. However, for the purposes of calculation, we keep the wing reference area fixed, so that any change in area is manifested within the change in the non-dimensional coefficient.

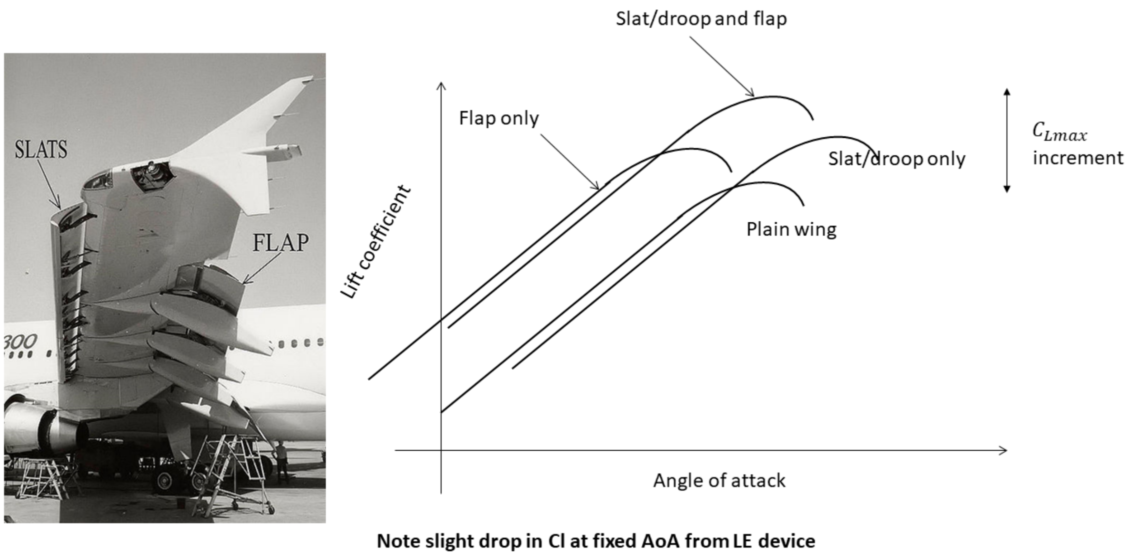


Figure 26: Influence of common high-lift devices on lift curve

Figure 27: Illustrations of types of high-lift devices

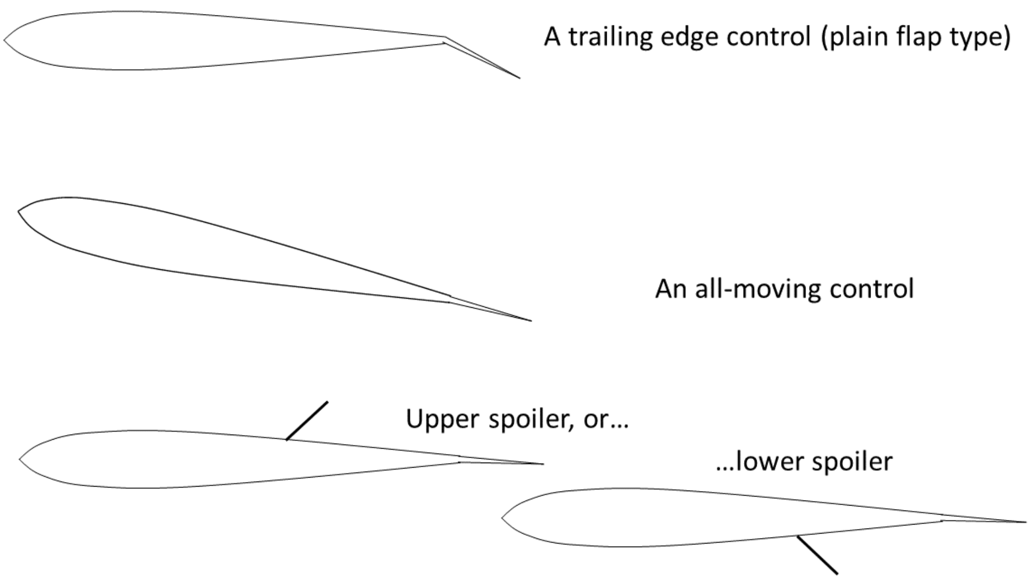


Figure 28: Control variations

5 Powered controls and Fly-by-wire

The maximum average pull strength of a male is about 400N, and for a female about 250N, directly overhead in a standing position. Seated, the maximum pull forces are similar, and the maximum push forces are around 230N for males and 100N for females.

So, male or female, seated or standing, humans are not that strong. What % of maximum would be reasonable as a control force to fly an aircraft for 8 hours?

To take a different approach,

$$H = C_H \frac{1}{2} \rho V^2 S_{control} c_{control} \tag{31}$$

For a hinge moment of 0.015 (which would be a deflection of about 10^o, hinged at leading edge), flying at 85m/s, with tailplane span of 10m and elevator chord 0.8m (for info tailplane chord about 3m), which roughly corresponds to a Lancaster elevator at cruise speed, we have

$$H = 0.015 \times \frac{1}{2} \times 1.225 \times 85^2 \times 10 \times 0.8 \times 0.8 \approx 424Nm \tag{32}$$

which is about 42kg at 1m, and exceeds maximum human pull. Without intervention, it would therefore not be possible to operate the control except for small angles, which might not be sufficient to manoeuvre.

Both women and men flew these aircraft. So, how was this resolved? Some mechanical advantage was used, but this was limited by cockpit space and not wanting excessive control movement (which would also be tiring). Prior to powered controls, which aircraft of this era did not use, there were two options: (i) move the hinge point of the elevator rearwards from the leading edge to reduce the aerodynamic moments about the pivot, or (ii) use a balance tab to providing an aerodynamic assisting force. Both approaches were used on aircraft, but primarily a balance tab reduced forces in this case. We'll look at these options now.

5.1 Aerodynamic balancing

From discussions regarding aerodynamic centre x_{ac} on aerofoils, we know this location is at about 25% of chord for an aerofoil in free air (wing of infinite aspect ratio). For a trailing edge control (ie. plain flap) the location of x_{ac} is not so obvious due to 3D variations and the influence of the upstream lifting surface (eg. tailplane or wing), but it will nonetheless be located towards the front edge of the surface. If you were to pivot the surface at this exact location, zero moment would be required to produce a deflection.

Completely balancing a control in this way would be quite dangerous, because it would lead to zero feedback on the controls. Pilots prefer some indication of high speed, and therefore the risk of high manoeuvre loads, to be conveyed via a 'stiffening' in the controls as speed increases. A fully balanced system would provide zero feedback, and would need to be held in position, but could still pitch, roll or yaw the aircraft violently, potentially exceeding design loads.

Therefore, a compromise is usually struck, with some - but not excessive - aero balancing. See for example the 'overhanging' balance in figure 29, which is an example of moving the pivot point.

A good example of aerodynamic balancing is on the elevator of the Spitfire, see figure 30, which uses a ‘horn’. In this case the surface is extended forward of the hinge line. Note that this extension also contains balance weights, which are intended to move the CG of the elevator forwards relative to the hinge, to help alleviate a risk of control flutter. So, the horn helps meet two requirements simultaneously.

5.1.1 Balance panels

A variant of moving the pivot, or adding area forward of the hinge, is the ‘balance panel’ in figure 29. These work in a very similar way, except they are not visually obvious. Rather, they are contained within the trailing edge of the wing/tailplane. Vents above and below mean that pressure is applied across the (sealed) panel. So, when the control moves down and a low pressure is generated above, this low pressure acts on the panel, located ahead of the hinge. This tends to produce a moment that rotates the control in the same direction, thereby assisting motion.

5.2 Trim, servo, spring and geared tabs

An aircraft will rarely be in equilibrium with zero force on the pilot’s controls. There may be asymmetries in yaw or roll due to propeller effects or fuel usage, but the most obvious feature is that changes in speed require changes in elevator position.

This is because if we wish to slow down, the angle of attack must be raised to compensate for the reduced dynamic pressure. The additional lift from the increase in angle of attack will act at the x_{ac} of the wing, which is behind the CG normally, therefore requiring a greater downforce on the tailplane, which is provided by nose-up elevator.

Unfortunately, it is uncomfortable if flying manually to hold the elevator in a particular position for long periods. You might want to scratch your nose, read a map or eat a biscuit (but not all at the same time perhaps). So, it’s important to have a way of adjusting the ‘zero force’ position of the elevator. To achieve this, elevators are equipped with trim tabs on their trailing edge. These are fixed relative to the elevator, and allow an aircraft to be trimmed to fly level at a particular speed without any force on the elevator control from the pilot. You could also trim an aircraft to fly a continuous turn, but this isn’t normally done, as turns are for short periods only compared to how long you may want to hold a different speed.

The interesting part is that this theme is not limited to trim. Tabs can also be used to provide aerodynamic force assistance to the pilot, if the tab is connected to the pilot’s controls instead of the elevator, which gives the *servo tab*. It may also be desirable to connect the elevator to the servo tab via a spring, giving the *spring tab*, or with a fixed ratio between them, giving the *geared spring tab*. You can also connect a tab to the fuselage, so that it will move if the whole tailplane pivots, as on the MD-80 series with the *anti-float tab*.

You can probably now appreciate that tabs of this type are (i) rather useful (ii) come in many variants and (iii) are a little tricky to get your head around. They are, however, quite important. The ATR turboprop series makes extensive use of spring tabs, while control redundancy on the 737 comes from servo tabs and balance panels. They may seem antiquated, but these systems remain on many modern aircraft and are likely to be an important part of control actuation on medium sized aircraft forever. The ancients knew some things, eh?

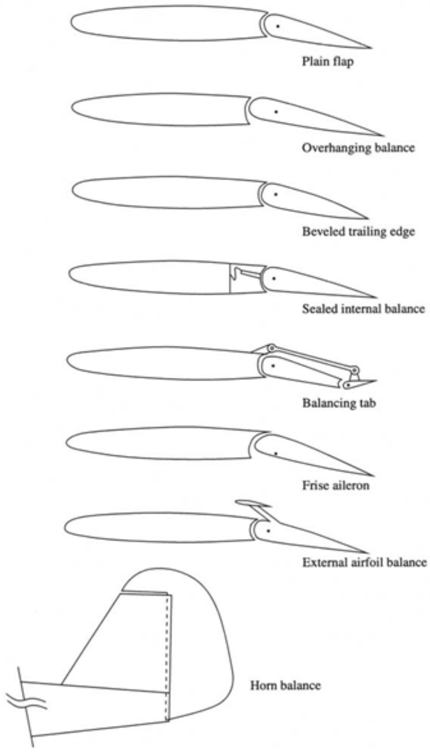


Figure 29: Variations on control balances

Figure 30: Horn aerodynamic balance on Spitfire elevator

As a footnote, larger aircraft use hydraulically or electrically powered controls. These give more rapid deflection and larger maximum deflection angles, which makes the controls of larger aircraft more responsive. It would be possible to fly them with servo tab systems, but the handling would be poorer, which is why they extra cost and complexity of these systems is worth it for larger aircraft.

5.3 Fly-by-wire

I don't really like the term 'fly-by-wire'. It would be better to describe it as 'get what you want' control, versus the old form of 'see what you get' control. It doesn't really have anything to do with electrical wires, for example; it's just *easier* to implement electrically!

So what do I mean? Well, on earlier aircraft, up to say 757/767 and A310, pilots moved the controls in the cockpit and the external flying controls (elevator, ailerons, rudder etc) moved in proportion to how far they moved the cockpit controls. The pilot then judged how much to move them manually - hence my 'see what you get' terminology.

FBW works on the basis that the pilot input is not a control deflection, but a *request to achieve a given outcome*. The pilot moves the sidestick to command a *roll rate*, not an aileron angle (why should the pilot have any interest in the aileron angle, anyway?). A computer then makes a rapid calculation, compares the current roll rate with the desired one, and works out how much to move the ailerons in order to achieve the desired roll rate. In pitch, normally the target variable is 'g' or normal acceleration, and in yaw usually yaw angle.

The controls themselves are therefore abstracted from the pilot, hence my term 'get what you want' control. Roll rate, 'g' and yaw angle are closer to what a pilot actually wants when they make an input, versus the angles of some controlled surfaces. Also, if no input is made and the aircraft flies through some rough air, the aircraft will automatically attempt to retain that zero condition with no pilot inputs.

Think about this - when you drive a car, what's more important - the angle of the front wheels, or how much left or right you are turning? You don't move the steering wheel to turn the wheels, you move it to turn the car at a given rate. The distinction becomes obvious if you compare parking to driving on a motorway. When parking, you want the steering wheel to turn the front wheels a lot, so you can squeeze in to that space without turning the wheel much. However, on the motorway, you want the front wheels to turn less for the same rotation of the steering wheel, so that the car is not so 'twitchy' holding position in a lane.

Clever car manufacturers now use electromechanical steering, which has a speed dependent steering sensitivity. My old car had mechanical-hydraulic fixed ratio steering, and it was more tiring on long motorway trips than my newer car which has variable ratio electromechanical steering (NB - 'new' here is a 10 year old car). I make fewer trips, too, because my passengers get tired of hearing me wax lyrical about it and choose to hitch a lift with someone else.

Anyway, FBW on aircraft is the same. There are so many variables you can use to improve how the aircraft controls behave - for example, speed, height, flap setting, thrust setting - that it is worth having a computerised system to fit it all together. At high speed, the controls move less for the same input, and at low speed they move more. The pilot just sees an aircraft that is 'comfortable' or 'easy' to control across the speed range. At the heart of the FBW approach is the inertial unit, which measures the roll rate, normal acceleration and yaw. This tells the control loop what the current values are, so that they may be compared to the commanded values to compute control deflection.

It doesn't stop there. Knowing all the inertial and air data, safety protection can also be implemented. For example, A320 implemented an 'envelope protection' system to prevent a stall occurring, by the system refusing further nose up elevator commands if the angle of attack (measured by a vane) was close to the stall angle. Rudder travel can also be used to prevent overloading the tailfin. The elevator command is limited so that the wing structure cannot be overloaded at high speeds. There are so many useful protections and convenient benefits to FBW that it is now standard in larger commercial aircraft.

It's worth pointing out the obvious, too. Contrary to popular belief FBW+envelope protection does not prevent an aircraft from crashing in to the ground - it just prevents it being in a stalled state when it does. Overall management of height and speed remain the responsibility of the pilot.

A further issue is what happens if information is lost, for example if the angle of attack vane is jammed by a bird impact, or the inertial unit fails. In these cases there are backup systems, and if those fail too, then the flight control laws will move to a degraded state with fewer protections. For example, if roll rate cannnot be measured due to a failure, the aircraft reverts from the standard 'normal' law to 'direct law', where pilot inputs move surfaces directly again. In between normal and direct, there are stages of 'alternate' law, where varying degrees of functionality are available depending on what exactly is broken.

I think some in the flying community feel FBW takes the 'feel' away from pilots. You can make your own mind up, but personally I don't think it does. Engineers *and* test pilots develop the FBW laws *together*, and providing operators understand how they work, they offer safer and more consistent handling across all aircraft speeds, configurations and payloads. Without a doubt, FBW and protection systems have by now prevented numerous accidents from occurring, and are here to stay.

6 Guidance, navigation and control

6.1 Guidance and navigation

It doesn't matter whether you're finding your way home in the evening or putting a rover on Mars, there are fundamentally only two ways to know where you are: (i) remember where you started, and the route you took, or (ii) recognise your surroundings in some way (I'll also agree a combination of those two is possible!). Both these approaches are used extensively in aerospace.

6.1.1 Celestial navigation

It may seem daft to discuss, but prior to the modern age navigation was already extremely accurate, as required for nautical purposes (landmarks being sufficient on land, except for example in deserts). The accuracy required to avoid submerged obstacles might be around a mile or so, and once near the coast sightings of landmarks would provide a better estimate, together with the use of *pilots* who had specific harbour knowledge (different to the pilots that fly aeroplanes). Most sinkings involved sailing ships unexpectedly caught on lee shores, rather than outright navigational errors.

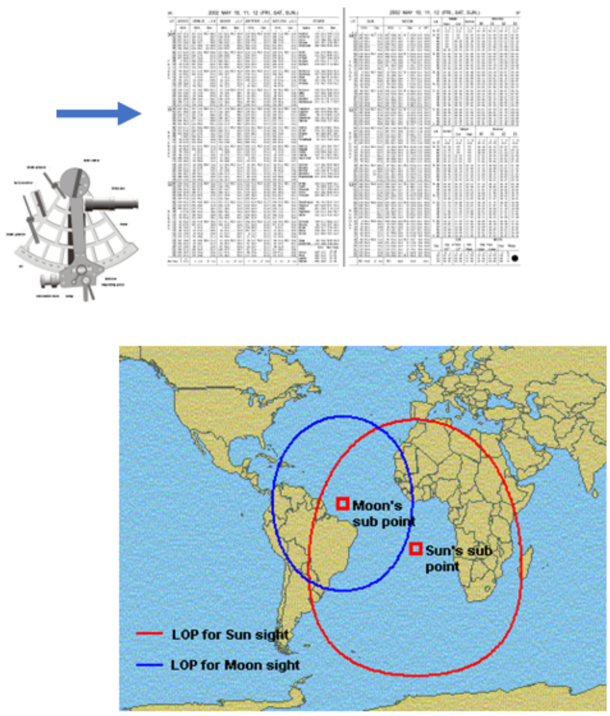


Figure 31: Celestial navigation

Early navigation involved measuring the angle of (typically) the sun to establish latitude (North/South) and then a 'chronometer' (watch) to find longitude. Later variants refined the use of the chronometer (known time is required to determine where the celestial bodies ought to be from tables).

So how is it done? It applies both the approaches (i) and (ii) above sequentially. The approach below is known as the 'intercept method' (developed by Captain Thomas Sumner in 1843 and Admiral Marcq Saint-Hilaire in 1875) - there were of course other variants on the way to developing this one.

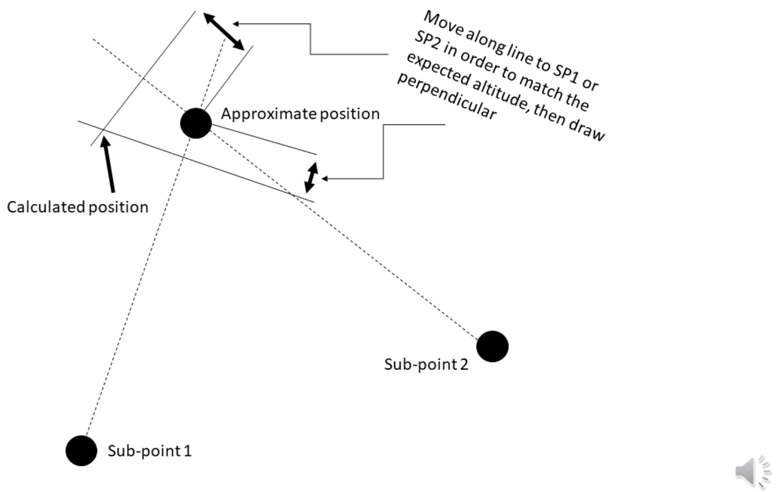


Figure 32: Celestial navigation ‘intercept method’

- Estimate you *approximate position*, or ‘AP’
- Observe the angle of known celestial bodies - stars, the moon etc., and using a sextant measure their angle above the horizon
- Using navigational tables, lookup the the expected locations of the celestial bodies on the Earth’s surface (‘sub-points’). This step requires knowing the time, and also your AP. Also look up the expected angle of the celestial body above the horizon
- Plot a line from your AP to each sub-point for each body
- Move along the line until the measured angle matches the one from the tables
- Plot another line at this point, essentially at 90 degrees to the previous one. These lines represent lines along which the measured angle of the body is constant, and equal to the one you found with the sextant. Technically these are not straight lines, but actually very large ovoidal loops - but the difference is small providing the AP is not drastically far from the real position
- Repeat for all bodies
- Where these two lines intersect is your calculated position. If there are three bodies, then you are at the triangle’s centroid, and for four, the centroid of the quadrilateral, and so on for more arbitrary polygons

This was, and in some ways still is, cutting edge stuff. Careful mathematics was needed to prepare the navigational tables, and rapid and correct use of the process needed on a ship to get the answer right quickly. Dropping the navigational tables over the side was also frowned upon, I believe.

6.1.2 Radio navigation

I think you can argue that radio navigation owes its history to making sightings of coastal geography, known as *pilotage*. With two measured bearings from known fixed features, and a map, the intersection is your position. Using more than one fixed feature will draw a polygon, with your likely position at the centroid. This process is known at multilateration, and as you can see by the previous discussion on celestial navigation, it is not a new concept.

It was presumably a natural navigational step to want to use fixed features you couldn't see, and from further away (hey, it might be foggy or dark, or you might be a long way from the shore). Two alternatives are possibles: either use radar to look for fixed landmarks (hills etc) beyond the horizon, or use radio transmissions to perform multilateration. The latter was more desirable (I think because if you look at a hill with a radar, it's hard to known which hill it is, especially if you're lost!). The option to use radio transmissions is known as *radio direction finding* and has a number of different implementations and variations relevant to aircraft.

The key point is how accurate you want your position to be. If you're landing with no visibility, you really want your position to within a few feet left/right and up/down; anything else and you risk an accident. If you're trying to maintain position at 10,000ft holding to land, position to within $\frac{1}{2}$ mile would be fine. If you're trying to catch a wire on an aircraft carrier, position and height should ideally be known to within inches. If you're performing aerial refuelling then it's a similar accuracy.

For all these needs there are variants of radio navigation. Measuring height accurately at low altitudes is perhaps one of the easiest, and most large aircraft carry radio altimeters that give height over the ground to the nearest foot, commonly heard as altitude callouts during landing. Manual landings with human eyeball judgement don't really require this (obviously many people land aircraft without radio altimeters) but it's helpful, especially in poor weather. Autopilots, however, definitely need a precise altitude to perform the landing flare correctly, as they have no vision data and the ILS (instrument landing system) glideslope isn't calibrated for each aircraft type, so it doesn't give actual wheel height above ground.

6.1.2.1 RDF With radio direction finding, known beacons transmit, and the receiver is rotated by the operator until the maximum signal direction is found, giving bearings to the separate beacons to allow multilateration. The rotating receiver often looks like a loop or similar.

6.1.2.2 Reverse RDF Large rotating receivers are a little unwieldy on aircraft, so RRDF was also developed. In this case, the transmitter sweeps a beam through 360 degrees, and the transmission includes information to determine when the beam pointed 'North', or along a reference direction. The receiver determines the timing of the maximum signal, and uses the time offset from the due 'North' time to work out the bearing to the transmitter. Multilateration is used again thereafter.

6.1.2.3 NDBs and VOR RDF technology underpins non-directional beacons, or 'NDBs', commonly used by aircraft. Reverse RDF underpins VOR (VHF omnidirectional range) systems. Both these systems remain in common usage, although there are fewer in operation now than there used to be due to the prevalence of satellite navigation. NDBs are cheaper than VORs, and the most expensive of all is a VOR with combined distance measuring equipment (DME), termed a VOR-DME.

NDBs and VORs mark airways, hold points and important track points on airport approaches. These points can also be located using satellite navigation, or 'GPS' (but GPS is only one variant, the US one, of the satellite navigation system). Unfortunately, it is unwise to put all the eggs in one basket, and if GPS were attacked or blocked, you could have a serious problem. For that reason, radio navigation is likely to remain for some time.

6.1.3 Satellite navigation

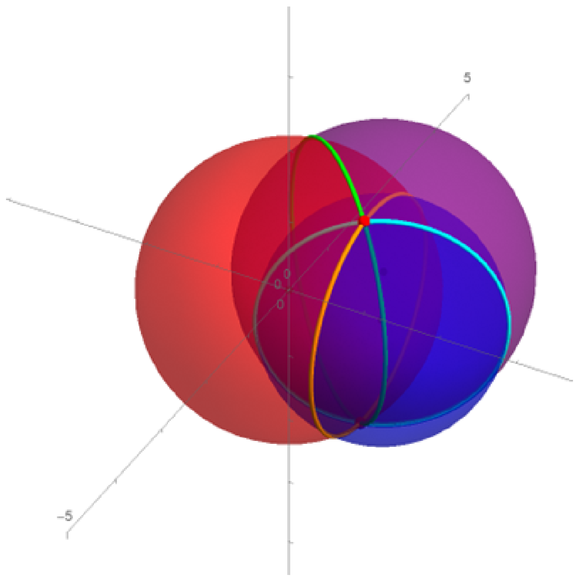


Figure 33: Some beautifully intersecting spheres illustrate how satellite navigation works

Truly the pièce d résistance of human ingenuity in navigation. If you can't see the stars, *put up some of your own*. It is essentially similar to RDF in 3D space, but using timing instead of signal strength, and falling in to the 'recognise your surroundings' category of navigation.

A certain number of satellites transmit a radio signal, which travels at the speed of light. The receiver knows *exact time* and is able to therefore calculate how far away the satellite is, which defines the receiver's position on a sphere in space. A second measurement provides another sphere, the intersection of which is circular. This circle probably only touches the Earth at one point, which is probably where you are, but in any case a third measurement will produce a single point.

Marvellous, eh? Yes. But also incredibly expensive - think about all those satellite launches. Nonetheless, amazing. But that's not the best bit. I said the receiver knew the *exact time*. The satellites know *exact time* because they have atomic clocks on board, so where is the atomic clock in your £200 telephone? It's not there, because something clever means it isn't needed, but knowledge of *exact time* still is, so how is it found?

The answer is that once you have 3+ measurements, you *have* to be at a single point, and as far as the system is concerned the only reason why you would not be at a single point is if you haven't measured the time exactly at the receiver's location. So, the process is to adjust the receiver's time *until a single location is produced*. Through this process, your phone achieves atomic clock level accuracy at no cost.

Wow! It's hard not to be amazed?

6.1.4 Inertial navigation

This is the ultimate in 'remember where you started and where you've been' navigation. The principle is so simple, written as finding velocity at some time t as the integral of all the accelerations measured from $t = 0$ up to that time t

$$\mathbf{v}(t) = \mathbf{v}_0 + \int_0^t \mathbf{a}(t')dt' \tag{33}$$

Then find position by integrating velocity

$$\mathbf{p}(t) = \mathbf{p}_0 + \int_0^t \mathbf{v}(t')dt' \tag{34}$$

so also

$$\mathbf{p}(t) = \mathbf{p}_0 + \int_0^t \int_0^{t''} \mathbf{a}(t')dt'dt'' \tag{35}$$

Accelerations can be measured by an inertial unit, often placed close to the centre of mass of a vehicle. It's incredibly good at measuring accelerations. Even so, there are two pitfall of an INS: you need accurate accelerations, and you also need a really accurate pair of $\mathbf{v}_0$ and $\mathbf{p}_0$.

If your accelerations are inaccurate, this is bad, because the error grows with time. The further you fly, the less accurately you know your location. Knowing $\mathbf{v}_0$ and $\mathbf{p}_0$ is easy, of course, as you usually know which airport you've left (not so easy if you're launching a system from a submarine under the ocean).

Current levels of accuracy mean that an INS might have an error of a few miles by the time you reach the west coast of the USA flying from Europe. So, it's not good enough for landing, but it's good enough to get you to the right suburb of the right city.

The most beautiful thing, of course, is that it doesn't require any ground based data. The worst thing is that it's not suitable for really long journeys where there are many varying accelerations; so, it's not suitable for ships, but it works well with aircraft.

A particularly effective variant is *astro-inertial* navigation, which is used on most ballistic missiles. In the event of their use, all land-based systems have probably been destroyed, and satellite navigation compromised. This means *only* inertial navigation and celestial navigation remain feasible. Particularly for submarine launched systems, it is important to remove the errors associated in $\mathbf{p}_0$ and $\mathbf{v}_0$, and this is done at the apogee of the ballistic flight by an automatic system that takes a star sighting to locate the missile relative to the star. With this information, an intermediate value of $\mathbf{p}$ is known, which substantially removes the accumulated integration errors. The believed accuracy of the Trident II system is $\approx$90m using this approach, which is sufficient for large targets. In fact, nearly all ballistic missile systems use this technique.

So, as it turns out, celestial navigation remains quite relevant, but perhaps it's a little sad that the end of the world would be computed using the same astro-navigational techniques that first led humans across the oceans?

6.2 Control

Control is the process of making any system do what you want it to do. Engineering loves to understand and model the original system, but once that is done, we usually want it to do a particular thing.

A simple example is central heating, which uses two stages of 'bang-bang' control. Let's say, when the temperature falls below the set value, your boiler lights and the central heating pump comes on.

When the water in the radiators reaches the boiler maximum temperature, the boiler turns off, but the pump continues to pump the hot water around the radiators until the room temperature reaches the set value. When the room temperature drops, the pump switches on again, and the boiler may also light again if the radiator water temperature has dropped.

Both the pump and the boiler can only be on or off; they are not continuously modulated (at least in most homes). This is a crude form of control, but it works in this scenario because your house temperature changes really slowly. This is an easy system to control. However, it will never succeed in maintaining an accurate temperature, because it will always fluctuate around the target value depending on the on/off state of the boiler and pump.

A more elaborate version is proportional control, which most of us have encountered in terms of cruise control in a car. The position of the accelerator is determined by the difference between target and desired speed, scaled by a constant, so

$$p_{accelerator} = k_p(v_{desired} - v_{actual}) \tag{36}$$

If desired exceeds actual, the accelerator position will increase. However, depending on k_p different things may happen. If k_p is big, very aggressive changes in accelerator position will be seen, and if it's small, the system will make almost no effort to change the speed, or it will take a long time. Aggressive changes in position will lead to overshooting the target speed and oscillations.

Furthermore, the system does not have a solution with $v_{desired} = v_{actual}$ because this would imply a zero accelerator position. Thus, the error may go to a small value, but never zero.

Rectifying this requires two further steps. The first is integral control

$$p_{accelerator} = k_i \int_0^t (v_{desired} - v_{actual})dt \tag{37}$$

while the second is derivative control

$$p_{accelerator} = k_d \frac{d(v_{desired} - v_{actual})}{dt} \tag{38}$$

Integral control fixes the issue of steady state error. The error integral grows with time, so the change in accelerator position will get bigger with time unless the target speed is met.

Derivative control says that if v_{actual} were increasing really quickly, then a negative k_d would be needed. This would reduce the accelerator position *before* the desired speed was reached, thereby reducing the likelihood of an overshoot.

Bringing these thee together gives PID control, which controls the vast majority of the world around us (erm, with the exception of your heating, probably, which is 'bang-bang', as discussed, or kettles and toasters...and early guided weapons).

6.2.1 Aerospace examples

Control for flight vehicles is more likely to target a given state. For example, you might used PID control to adjust an aircraft's elevator until the climb rate matched the autopilot's selected value. Or,

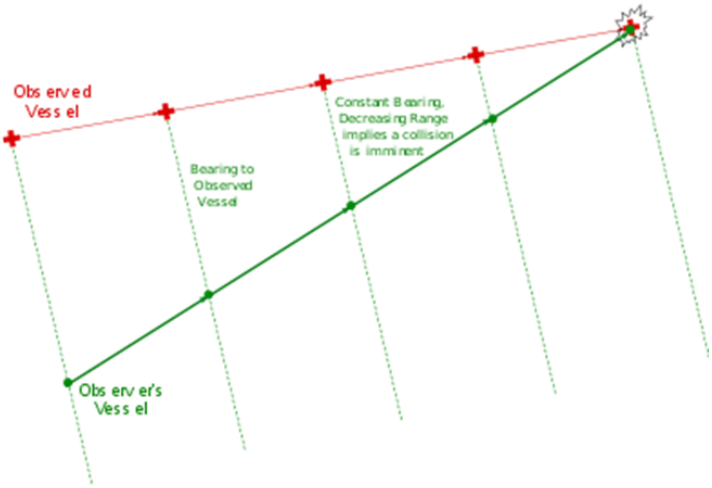


Figure 34: The ‘constant bearing, decreasing range’ principle

you might control the ailerons until the aircraft’s heading met a fixed value, together with control of the elevators to target a fixed altitude and keep the aircraft at the same altitude (this would ensure a balanced turn).

A comparative example of control comes from the Paveway I and II guided bombs. These early variants in fact used ‘bang-bang’ control in a similar way to your home heating system to track the target, with control fins moving 100% as soon as the threshold was reached. Wiggly flight paths and associated reduced range due to higher drag on the deflected fins led to adoption of PID control in later variants. Generally, navigation of these types of systems is a combination of INS, GPS and optical designation, depending on prevailing weather conditions.

A further example is the AIM-9 sidewinder heat-seeking air-to-air missile. This is a particular example of *proportional navigation* used for pursuit a target. The system uses rotating mirror to record target position, and controls its own trajectory (using PID control) *until the position of the target relative to the missile remains exactly constant*. This is also known as a ‘constant bearing decreasing range’ strategy, and if followed guarantees a collision, as shown in figure 34. But why? And why is this better than pointing the nose of the missile at the target (‘direct pursuit’)?

The answer is that this strategy forces the missile to ‘lead’ the target: instead of aiming at it, it aims at where it will be in the future. Consider driving a car towards a pedestrian crossing. If a person is walking towards the crossing, they may be behind the windscreen pillar of your car. If you drive forwards at the same time as they walk towards the crossing, at just the right rate, they will remain exactly behind the pillar. Furthermore, if this continues, you will hit them with the corner of your car that lines up with a vector from your eyes directly towards the pillar. This is a considerable risk when driving.

It’s also a risk when flying. If you observe another aircraft, and it is moving relative to you, you are not going to hit it. If you observe it twice, and it has not moved between observations, you are currently on a collision course (a simple example is an aircraft directly ahead, coming in the opposite direction). This probably explains in some way many collisions, because our eyes look for things that are moving relative to us. Unfortunately, it is the constant and not the variable that is a hazard in this situation. It probably also explains those occasions when you bump in to someone ‘because you weren’t looking where you were going’. Actually, you probably were looking where you were going, but didn’t notice something that hadn’t changed. It really is a cruel combination of psychology and maths.

But wait a minute, how does a sidewinder avoid rolling all the time? Unless the roll is constant (ie. near zero roll rate), proportional navigation won’t work, because the reference orientation will keep changing. The answer - gyroscopically driven aerodynamically powered ‘rollerons’ on the fins stabilise the missile in roll and prevent high roll rates entirely.

7 Propulsion

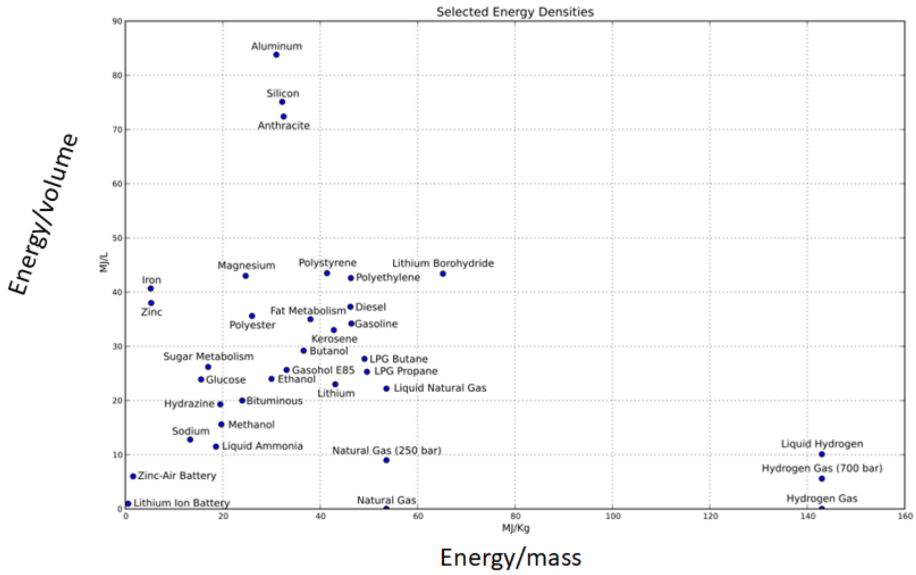


Figure 35: Properties of different fuel types

One of the staff who taught me, back when dinosaurs roamed the the lecture theatres, stated that an aircraft without engines is merely a ‘caravan with wings’. It’s a fair point. Forward motion is needed for lift and covering distance, so assuming we’re not going to try and ride the thermals to New York, we need propulsion.

7.1 General points

7.1.1 Combustion

We’ll start by covering air-breathing engines. So, first of all, what’s combustion? Let’s look at combustion of kerosene/Diesel (NB Diesel is capitalised after Rudolf Diesel, the fascinating and tragic creator of the engine bearing his name, however, there was no Mr Petrol, hence petrol (short for petroleum))

$$2C_{12}H_{26} + 37O_2 \rightarrow 24CO_2 + 26H_2O \tag{39}$$

The equation is ‘balanced’ in chemical terminology; there are 24 carbon atoms left and right, 52 hydrogen and 74 oxygen, and the reaction is of course exothermic (releasing energy). It’s worth pointing out that other combustion reactions are possible if there is not full and ready availability of oxygen, known as ‘incomplete combustion’. Think of when you close the collar on a Bunsen flame and it becomes yellow and sooty, which is characteristic of incomplete combustion. Incomplete combustion is (i) wasteful and (ii) dangerous, because it leads to poisonous by-products like CO (carbon monoxide). It’s a good reason to never leave your barbeque in your tent, and to be very worried if you see sooty black deposits around your boiler in your home (if you sleep in a room with a burning appliance, you should always buy a carbon monoxide detector). But I digress.

Contrary to popular belief, complete combustion isn’t hazardous (neither carbon dioxide nor water pose a risk to us, except via displacing oxygen or climate change). However, it’s incredibly tricky to get proper mixing everywhere meaning that even if you're careful, some incomplete combustion will accompany it, generating less-than-desirable or outright poisonous by-products. Also, at high temperatures, nitrogen and oxygen will always react together to produce nitrous oxides, which are harmful to humans too. Both jet and Diesel engines are thermodynamically more efficient the hotter they burn (as is any engine, in fact), so you can perhaps see the conundrum. The difference is that jet engines leave the NO high in the sky, whereas motor vehicles leave it at street level.

Anyway, the beauty of the balanced chemical equation is that we can work out how much of one thing reacts with another. We know that 2 moles of kerosene needs 37 moles of oxygen, where a mole is $6.02214076 \times 10^{23}$ molecules (check out the definition of Avogadro's number if you want to know more). A mole of kerosene is 170g, and a mole of oxygen is 32g, so we need $\frac{37}{2} \times 32 = 592g$ of oxygen, which is 2.96kg of air (at about 20% oxygen), giving a fuel-air ratio of 17.4 by mass. Each kg of fuel produces around 3.1kg of CO_2, and a large aircraft might use 150 tonnes of fuel, so it's an impressive amount of emissions in total. The white vapour behind an aircraft is almost exclusively gaseous water condensing in the cold air at altitude, also dependent on the humidity at altitude. Water is harmless, except that condensation trails - contrails - can lead to clouds which may also influence the Earth's reflected heat. Quite often you also see water dripping from vehicle exhausts, and condensing boilers do it to warm your bath a bit more efficiently. It's all just part of burning hydrocarbons (I always find it weird that if you burn methane gas, you'll end up with some water, but hey that's chemistry).

However, kerosene is not a perfect 100% blend of one hydrocarbon, so the 'stoichiometric' ratio of 37/2 is not actually quite exact. Other reasons, including reducing emissions, mean you might not run on the exact ratio, so in practice kerosene/Diesel runs around 15-16. Petrol is around 14.7.

Most everyday combustion is somewhat incomplete. Candles, open fires and so on all look yellow and sooty. Gas hobs are good, because natural gas mixes well with air in your kitchen and so there's enough all round oxygen to balance the equation well (also, if they didn't we would forever be poisoning ourselves at dinner time with CO). Wax is a long chain hydrocarbon and you would need to work hard to make sure it sees enough oxygen to react fully. If the airport fuel depot catches fire, it will be smoky, because the liquid fuel will not be fully reacted, whereas in a jet engine it nearly will be and hence there is little smoke from a jet. Unless of course you go back to the bad ol' days of turbojets, which had to be pushed up to and beyond stoichiometric ratios to get takeoff thrust, including use of water injection. Turbofans offer much more efficient low-speed power, so you shouldn't see an A350 or 787 doing a smoky takeoff.

The combustion chamber of a jet it essentially the design solution for complete combustion. Fuel sprays in at a carefully regulated rate and combusts in the presence of the right amount of oxygen. Dispersal means no droplet of fuel is prevented from seeing enough oxygen by its neighbouring droplets.

Ok, so that's combustion, why do we need to compress the air?

7.1.2 Compression

Compression achieves two things: (i) there's more oxygen in a smaller volume, so engine power can be raised (ie. you increase fuel flow rate to match the oxygen flow rate) and (ii) having the gas at high pressure means that once heat has been added, expansion can be used to extract mechanical work, either through a turbine or via a piston. The fact that gas gets hotter when it is compressed is somewhat undesirable, as the temperature increase lowers density (and thus quantity of oxygen in a finite space), which is why some internal combustion engines use intercoolers, like the one on my old VW, which always used to drip oil. Jet engines do not use coolers unless forward speed is very high (see the Skylon engine design, for example). Intercoolers at lower speed would be too heavy and bulky (bulky engines generate drag).

So, all things done, higher compression is a good thing, because it means more expansion can be done later to extact more mechanical work. Unfortunately, the more compression you want, the longer, wider and heavier your engine becomes because the compressor needs a greater mass flow and needs more stages to achieve the higher compression (if you try to increase pressure too much across a compressor stage, the blades will stall). Thus, there is a compromise between high compression for fuel economy and low compression to make a small, light, low-drag engine. Isn't engineering fascinating?

As a footnote, compression of a piston engine is limited by the forces on the mechanical parts. Jets can always increase compression by adding more compressor stages, if it's an axial compressor, anyway.

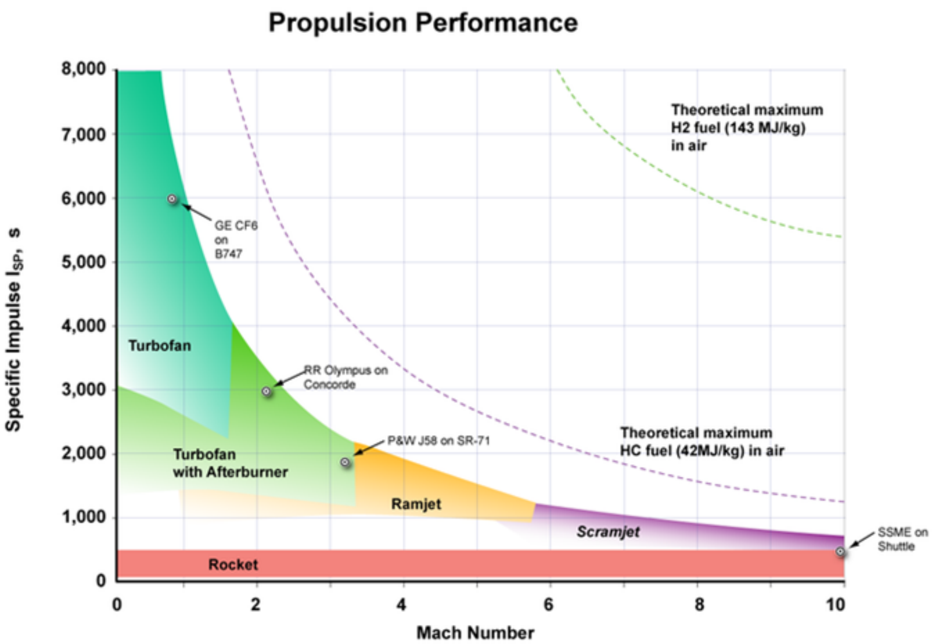


Figure 36: Properties of different fuel types

7.2 Analysis

Let's define specific impulse SI

$$I = \frac{Tt}{mg} = \frac{T}{\dot{m}g} \tag{40}$$

which is the thrust T per weight flow rate of fuel, and has dimensions of just s^{-1}. You might also meet *thrust specific fuel consumption*, which is

$$TSFC = \frac{\dot{m}}{T} \tag{41}$$

ie. the reciprocal of SI, without the g, and is the thrust per unit mass flow rate of fuel.

Can we do something cunning to work out how I changes with forward speed? Indeed we can. Let's assume the power generated by combustion all goes in to propulsive power (such that thermal and propulsive efficiencues are exactly one, see later), so that

$$P = Tu_0 \tag{42}$$

where u_0 is forward speed and T thrust. Then

$$I = \frac{\frac{P}{u_0}}{\dot{m}g} = \frac{h}{u_0 g} \tag{43}$$

where h is enthaply of combustion of the fuel, ie. the Joules released by combustion of 1kg of fuel. There's not much you can do to change h, and for kerosene it is about 43.1MJ/kg. So, using this data you can work out what the best SI is for a kerosene engine at a particular speed. For liquid hydrogen, h is about 144MJ/kg. Fairly obviously, liquid hydrogen is better on an energy per unit mass level (but it's a pain to store, because it's really cold!).

Did I tell you about rockets? The analysis is straightforward here too. The mass of 'air' seeing the momentum change is just the fuel and oxidiser together, which is the total fuel. The exit speed from the rocket is v_e and the fuel starts at zero velocity in the rocket's frame of reference, so

$$I_{rocket} = \frac{\dot{m}_{air}\Delta u}{\dot{m}g} = \frac{v_e}{g} \tag{44}$$

which means rockets seem to maximise v_e, which is achieved through the nozzle. Note that there is a further thrust contribution from pressure thrust if a nozzle if supersonic, but that's for year 2. Interestingly, an ideal nozzle doesn't do that anyway, so it is a secondary point. Hydrogen rockets don't exceed an SI of 500s or so, because the optimal ratio of chemical energy to mass of water to use to impart momentum is fixed by the chemistry. Nuclear rockets, or those using different fuels, can achieve different values of course.

7.3 Propulsive efficiencies

Let's break the efficiency in to parts

$$\eta_{total} = \eta_{thermal}\eta_{propulsive} = \frac{\text{rate kinetic energy}}{\text{fuel power}}\frac{\text{propulsive power}}{\text{rate kinetic energy}} \tag{45}$$

$$= \frac{\frac{1}{2}\left(\dot{m}_e u_e^2 - \dot{m}_o u_o^2\right)}{\dot{m}_f h}\frac{Tu_0}{\frac{1}{2}\left(\dot{m}_e u_e^2 - \dot{m}_o u_o^2\right)} \tag{46}$$

where $\dot{m}_e$ is exit mass flow rate (including fuel mass), $\dot{m}_o$ is the (air) inflow mass flow rate, h is enthalpy of combustion, T is thrust and u speed (either at inflow or exit).

Fuel mass added is small (remember the ratios above from the chemistry), so replace with m only and then look at propulsive efficiency only

$$T = \dot{m}(u_e - u_0) = \dot{m}u_0\left(\frac{u_e}{u_0} - 1\right) \tag{47}$$

$$\eta_{propulsive} = \frac{\dot{m}u_0(u_e - u_0)}{\frac{1}{2}\dot{m}(u_e - u_0)(u_e + u_0)} = \frac{2u_0}{u_e + u_e} = \frac{2}{1 + \frac{u_e}{u_0}} \tag{48}$$

There are two different limit situations that can be explored

1. high $\frac{u_e}{u_0}$, high $\frac{T}{\dot{m}u_0}$, low $\eta_{propulsive}$, lower drag smaller/lighter engine/intake
2. low $\frac{u_e}{u_0}$, low $\frac{T}{\dot{m}u_0}$, high $\eta_{propulsive}$, higher drag larger/heavier engine/intake

It is therefore not clear whether one ought to pursue higher propulsive efficiencies; it really depends on the size and weight of the engine that would be produced. This is essentially the compromise that sets the optimal bypass ratio for commercial turbofans, together with more prosaic issues like fuel and maintenance costs.

In turn, this brings us to the most interesting outcomes of this analysis, which is to derive the fuel usage for turbofans and turboprops. In fact, the idea is more general than that, and independent of the engine architecture. It only depends on whether your engine has high or low propulsive efficiency.

$$P = \frac{1}{2}\rho A u_0(u_e^2 - u_0^2) = \frac{1}{2}\rho A u_0(u_e + u_0)(u_e - u_0) = \frac{1}{2}T(u_e + u_0) \tag{49}$$

Now, if propulsive efficiency is very high, then $u_e \approx u_0$ and $P = Tu_0$. That is, fuel usage will be proportional to *power*. This is the case for propellers, which have good propulsive efficiency. It doesn't really matter what powers the propeller - it could be a turboprop, or a piston engine.

If propulsive efficiency is very low, then $P \approx \frac{Tu_e}{2}$ and fuel usage is proportional to *thrust*. This is the case for turbojets at low speed. Again, it doesn't matter what powers the engine specifically, merely that propulsive efficiency is lower.

Turbofans have fuel usage that is both power and thrust dependent. Usually, though, they are characterised as showing thrust dependent usage. Turbojets have terrible propulsive efficiency at low speeds, and are definitely thrust dependent in fuel usage. Note, though, that their propulsive efficiency improves with speed as a result of this.

Why should we care about this? Well, it is critical in determining optimal operating points for turbofans and turbojets. You will see this in the aeronautics material; power and thrust specific fuel consumption makes a big difference in where you fly in terms of Mach number. Without this analysis, it's a mysterious concept, but hopefully this has helped make it clearer. It is essentially why we see turbofans operated at transonic speeds, and turboprops at below transonic conditions.

7.4 Engine types

The mantra of suck-squeeze-bang-blow goes a long way to describing combustion engines. It might however be sensible to replace 'bang' with 'heat', as there is often no distinct explosive phase apart from in the internal combustion (piston) engine.

7.4.1 Internal combustion

Let's describe the 'four stroke cycle', so called because there are four strokes and stages before the cycle repeats.

The piston moves down, drawing a mixture of fuel and air in to the cylinder. The piston then moves up again, compressing the mixture, until it ignites. The increase in pressure from the ignition and rise in temperature forces the piston back down again. When it comes up again, it pushes out the burnt gases.

Inflow and outflow of air/fuel mix and exhaust gases it managed by valves, usually driven by cams and springs. Ignition is triggered by a spark in the petrol engine, but in the Diesel engine combustion starts due to the intense compression stroke raising the temperature (when first started, glow plugs warm the cylinder head to start things off). Therefore, the only thing controlling the speed of a Diesel is the amount of fuel flowing, so long as air is readily available. Due to the higher pressures, the fuel is often injected, which also helps with precise adjustment of combustion in the cylinder.

It's worth mentioning the 'two stroke' cycle. It's surprising this works, but it does. As the piston moves down, exhaust gases are expelled just before the mix is drawn in, by a cunning arrangement of the inlet/outlet ports (there may in fact be no valves). There is then a compression stroke, followed by the cycle completing.

Two stoke engines offer higher power to weight ratios, cheap manufacturing, poorer fuel economy, greater vibration and poorer emissions. You will find them on garden equipment and scooters/small motorbikes. Two strokes can also use compression-ignition, similar to Diesel engines, as seen on model aircraft engines.

Of course, there's a lot more to it than that, but internal combustion engines are not fundamentally complicated.

7.4.2 Turbojet

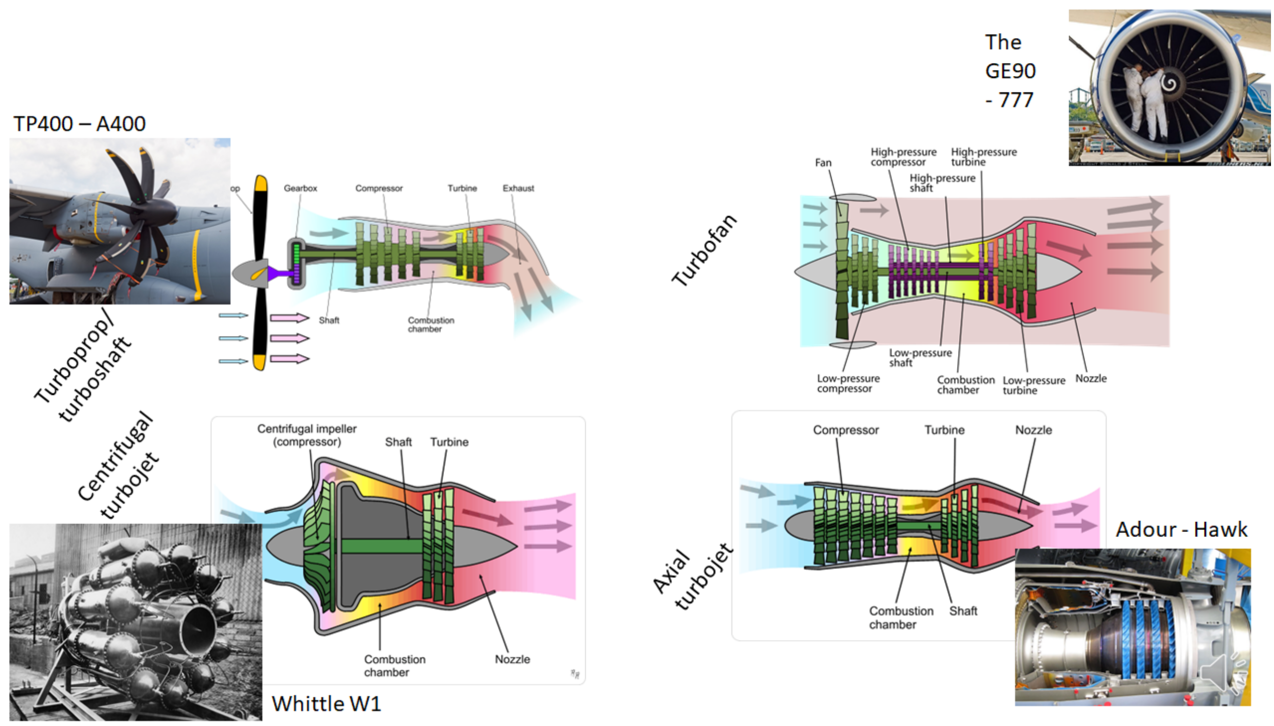


Figure 37: Types of jet engine

If you thought piston engines were too complicated then you will be pleased to hear the jet is simpler, almost unbelievably. We take some air, compress it with ‘something that spins’ (a compressor), add heat and then let it expand, also over ‘something that spins’ (a turbine).

So where did the idea come from? Technically, Hans von Ohain (Germany) and Frank Whittle (UK), but let’s dig deeper than that, because they obviously got their ideas from other sources too.

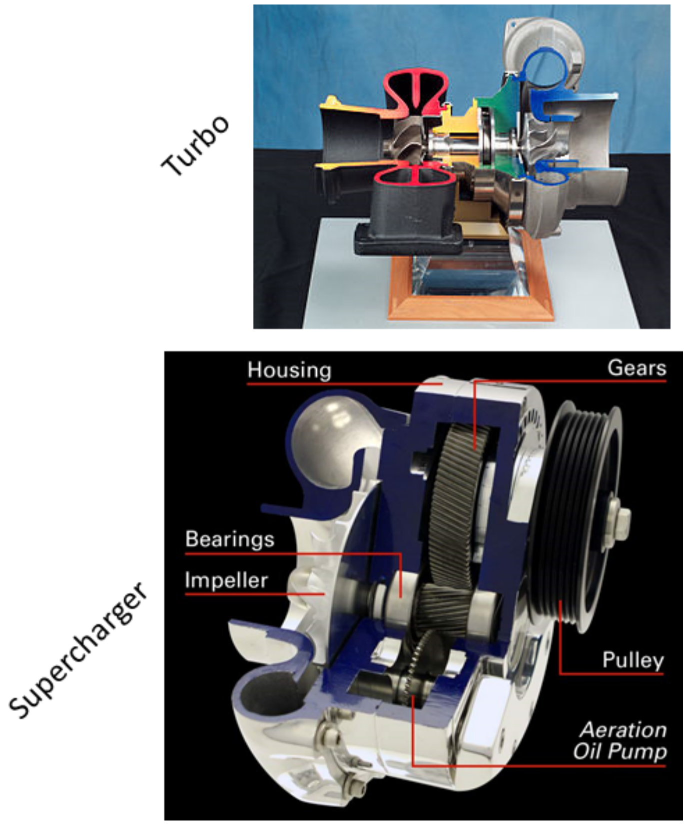


Figure 38: A turbocharger and a supercharger

For some time both superchargers and turbochargers (figure 38) had been used on piston engines. The turbocharger uses moving exhaust gases to drive a turbine, on the other end of which is a centrifugal compressor. The compressor ‘charges’ the cylinder with air at a higher pressure/density, allowing more fuel to be burnt stoichiometrically. Often there is an ‘intercooler’, which cools the compressed air to raise its density further.

Superchargers are the same, except there is no turbine, and a mechanical belt or shaft drives the compressor. Many car engines, especially small petrol models, are now both turbo and supercharged to improve fuel economy and power/torgue per unit weight across the whole RPM range.

So, Whittle and von Ohain were aware of this. They were also aware that steam turbines had been powering ships for a long time, in addition to the more conventional piston-based steam engines.

It was therefore not a wild idea to believe that the compressor could be driven by a turbine (because thats how a turbo works anyway), but with the air from the compressor passing over the turbine after combustion. What *was* a wild idea was to think it was possible to make such a device, without the heat destroying the turbine blades.

So, to make a jet engine, you take a car turbocharger and duct the air in to a chamber, where you add fuel and combust the mix. The hot gases then need to be ducted over the turbine, which in turn connects to the compressor, as they are on opposite ends of the same shaft. You get noise, heat, smelly gases, and of course, a reaction force from expelling the gases at high speed.

This is the turbojet.

Of course, there are one or two details. The turbine metal gets very hot in the combustion stream, and to begin with didn't last long. The parts are small and hard to machine accurately. The combustion chamber is intensely hot and welds there easily fail. Ultimately, your engine needs to be light, too, to drive an aircraft.

So there was no shortage of challenges that Frank Whittle and Hans von Ohain had to overcome. If you want to know more, go and read *Genesis of the Jet*. As it turns out, the story is more about people, their beliefs, prejudices and fears, and government funding, than it is about physics. Whittle and von Ohain probably always knew they were right, because the germ of their idea was simple and made physical sense.

7.4.3 Turbofan

It certainly doesn't end with the turbojet though. The worst turbojets (like the ones you are I might make) probably only just sustain themselves in terms of power balance. They are noisy and inconvenient fireplaces. They can be improved, and work quite efficiently at high speeds, but for transonic flight they are not fuel economic.

The reason for this is that the stoichiometric combustion of air produces a certain amount of energy, which produces a certain amount of kinetic energy. If you try and extract more of this kinetic energy as work, the engine will be bulky and heavy. You are tied to the fact that all the air passing through the engine gets combusted, and there's nothing that can change that. The exhaust speed is therefore high, and much kinetic energy is wasted.

But wait, why is all the air combusted? Why not...introduce alternative gas streams?

Thus the turbofan was born, a well as all the other jet variants. Shooting a hot meteor out of the back to produce thrust is wasteful, so instead put another compressor/fan/ducted fan on the front. The combustion core of the engine is now much smaller than the fan radius, and a lot more air goes past the core than through it. The exhaust speed from the fan is set by its blade pitch and RPM, not by stoichiometry of air and kerosene. The core and the fan are *nearly* decoupled, but not quite, because the final turbine stage is what drives the fan.

7.4.4 Turborop

Let's expand on the turbofan idea. The turbofan has a duct on the front which reduces the forward airspeed the fan sees from Mach 0.8 to Mach 0.3 to 0.4. Without the duct, the fan blades would be stalled at higher speeds.

Remove the duct and then allow the angle of the blades to be changed, and you have a turboprop - a turbine powered propeller. The bonus is - no duct, and also an even higher bypass ratio - most of the air goes through the propeller, not the engine core. This gives even better propulsive efficiency at lower speed.

Of course, at higher speed the propeller cannot increase the speed of the air by enough to generate much thrust. Also, as the RPM of the propeller rises to match the forward speed, the blade tips become supersonic, which is unbelievably unpleasant inside the aircraft. There is therefore a speed limit on turboprops in terms of capability and comfort.

Nonetheless, they are economical, and can have long ranges (although most are nowadays quite small). BOAC used Bristol Britannias on their long range routes for a period of time, and the Tu-95 remains in service as a long range bomber. Turboprops in commercial use today are usually short range ATR-42/72s or Dash-8 Q400s. The reason is that time is money, and the fixed costs, like crew wages, favour faster aircraft on long routes because the same aircraft can generate more seat-miles if it is quicker, while the crew cost is the same. Faster aircraft are less effective on short routes because the time in the cruise is a smaller fraction of the flight.

It's just possible that if fuel costs become extremely high relative to other operating costs we might see long range turboprops again, but I doubt if customers would accept the longer flight times, and the fuel savings would have to balance out the reduced production of seat-miles (due to slower flights).

7.4.5 Turboshaft

Can we go further with bypass ratio than a turboprop? Certainly, we can.

Helicopters are awkward machines that live on the edge of feasible power and efficiency. They cannot afford to waste a gram of mass or a Watt of power. The downwards velocity induced through their rotors has to be really quite small, as wasting kinetic energy in the air to produce thrust will degrade their already marginal performance. Nonetheless, there is an engine to suit them.

Instead of driving a propeller, a *turboshaft* drives a rotor through a rather complicated gearbox (the RPM of the rotor is much lower than the propeller on a turboprop, but the RPM of the turbine RPM is similar in both cases, hence a larger gearbox is needed on a turboshaft).

There some is redemption for helicopters in the form of tilt wings and tilt rotors, which also use turboshaft designs. Helicopters are one-trick ponies, but it is a popular trick (taking off vertically).

Of course there's one thing worse than a helicopter for power usage and efficiency, and that's a vertical takeoff or landing fast jet, like the Harrier or F-35. The exit velocity from the Harrier's nozzles is so high that the hover efficiency is terrible. The F-35 tries to improve this by using a lift-fan to get a better bypass ratio in the hover, allowing a higher payload. The Harrier's hover time is quite limited (in fuel and engine temperature/wear) and the F-35 can always carry more taking off from a runway.

7.4.6 Anything else?

My favourite conversation piece to test if people think I'm deranged is to discuss nuclear jet propulsion. It's interesting because people think it is a remote or impossible idea, when in fact it has been tested many years ago (project *Pluto* as well as the AN programme which lead to two nuclear powered turbojets), and conceivably remains feasible.

Look at all the jet engines above. The heat comes from burning fuel. That is the only thing that links them to kerosene. Otherwise, they are merely tools that turn hot gas in to thrust in some form.

If you were to heat the air in another way, they would still work just the same. If you put a reactor core in there, either directly or via a heat exchanger, you could build nuclear turbojets, turbofans, turboprops and turboshafts. There are no physical principles to preclude it, only the difficulties of miniaturisation.

As it happens, the probability is that these systems are actually in operation on some long distance weapons. A nuclear jet aircraft would look and sound much like a kerosene fuelled one. The only exterior difference would be an absence of smelly fumes, and all the concerned faces (!).

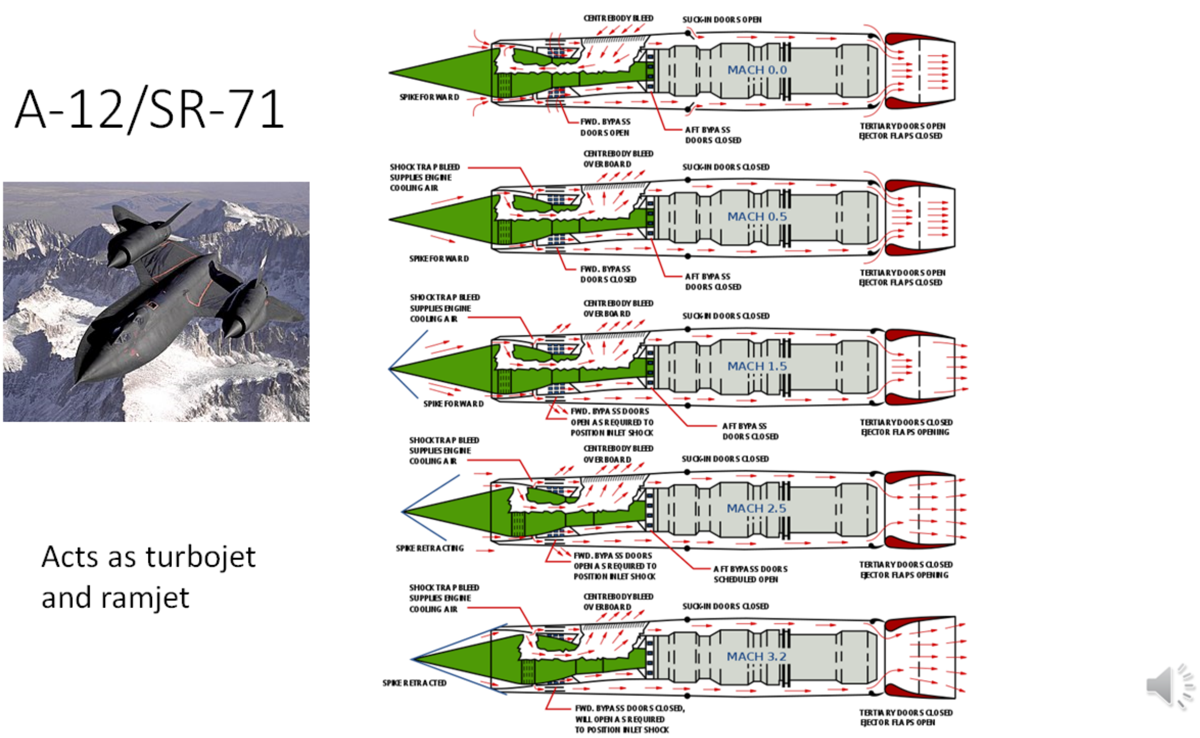


Figure 39: Cycle variants of the SR-71 engine

8 Civil aircraft design

Let's talk about what it is we want a commercial aircraft to do. Self-evidently, the purpose must be to make money, because they are *commercial* aircraft, not recreational aircraft. We might love them for their technical prowess, but airlines see them as cash cows. Airlines don't have an interest in aircraft, fuel, crew or maintenance; their priority is to make a return on a capital investment, the same as any other corporation, or indeed the whole economy.

It gets more interesting when we try to map how to make money to how to operate an aircraft. This is intensely difficult and depends on how much you pay your crew, what you pay for fuel and maintenance and what price you can sell your tickets for. The truth is, it's airline specific, and therefore manufacturers do simplify the objective a little.

The most obvious goal, in car parlance, is to maximise miles-per-gallon. We can explore this by looking at Breguet range for a thrust specific (eg. jet) and a power specific (eg. turboprop) (refer to aeronautics notes for derivation)

$$R_{\text{thrust specific}} = \frac{a}{gc_T}\frac{ML}{D}\ln\left(\frac{W_1}{W_2}\right) \tag{50}$$

$$R_{\text{power specific}} = \frac{\eta_p}{gc_P}\frac{L}{D}\ln\left(\frac{W_1}{W_2}\right) \tag{51}$$

The obvious outcome is that thrust specific turbofans/turbojets improve range at higher Mach number M. In comparison, a power specific aircraft such as a turboprop seeks to improve the propulsive efficiency of the propeller rather than increasing M. It is hard to be certain, but in general, one might expect the best turboprop to beat the best turbofan. This doesn't seem to have happened in the real world, because something else matters just as much as best miles-per-gallon.

Linking back to the discussion on turboprops, let's say it costs the same to fly a certain aircraft whether it is a turboprop or turbofan in terms of crew, maintenance, depreciation and interest on borrowing. So, the only difference is that one flies faster, burning more fuel. Unless fuel costs are very high, the faster aircraft is preferable because is generates seat-miles (the saleable quantity) at a higher rate. Thus, turbofans are preferable.

Conversely, Concorde took this a little too far. Although the Mach number was high L/D was low due to supersonic flight, so fuel usage was costly. A turbofan aircraft was economically preferable, although Concorde was a wonderful technical achievement. It wasn't just fuel though; maintenance was also a high cost on the aircraft.

On the other hand, if an aircraft spends little time at its optimal ML/D, and more time climbing/descending, then cruise performance is a moot point. This drives the preferance back to turboprops for short flights. You can see how ML/D varies in figure 40; it goes up in proportion to M, but then drops off after a maximum due to increases in wave drag at transonic speed.

Even when it seems like speed shouldn't matter, as for the distance-record-holding turbofan-powered *GlobalFlyer* (which flew around the world and then across the Atlantic again in about 77hrs), it sometimes still does. Humans only have so many hours of useful wakefulness, so speed still mattered to get as far as possible before Steve Fossett became exhausted. Nonetheless, the previous record holder for range was *Voyager* flying around the world in 216hrs, but with a crew of two, and Voyager

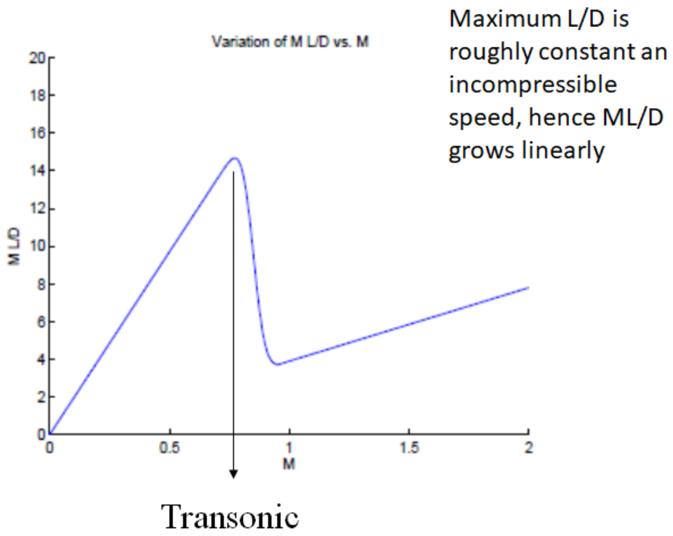


Figure 40: Sketch of ML/D variation with M

was a piston-engined propeller-driven aircraft. If you think those records are impressive, consider also *The Spirit of Butts Farm*, a piston model aircraft that crossed the Atlantic in 2003 in 39hrs. It may perhaps be that props offer the best possible endurance, and turbofans the best possible range.

8.1 Operating point selection in terms of $\frac{ML}{D}$ or $\frac{\eta_p L}{D}$

First consider the operation of a turbofan (thrust specific engine) and look to maximise $\frac{ML}{D}$, which is aerodynamically more interesting than the $\frac{\eta_p L}{D}$ of a turboprop (for the reasons we're about to cover!).

The highest $\frac{L}{D}$ that an aircraft can achieve is approximately constant until we reach the transonic Mach number range. There is probably a *slight* improvement in $\frac{L}{D}$ as speed increases because C_D will drop as Re increases, but this is a small effect. Fairly obviously then, ML/D will grow linearly with M.

This doesn't continue forever, though, because as some value the critical Mach number M_{crit} (typically 0.7 to 0.75), the air at some location on the wing will become supersonic. Once this happens there will be normal shockwaves at the end of the supersonic region, and these will cause drag (i) directly via momentum imparted to the air and (ii) indirectly via boundary layer thickening (which is acceptable) or separation (which would normally be avoided by design). This means that the best $\frac{L}{D}$ starts to fall as M rises, and, therefore, there is a maximum in ML/D at some point. It also means that for some fixed M, there is an optimal C_L for best L/D. You can see these curves in figure 41.

The operating point is hence made up of two values - the optimal M and the optimal C_L, which together give the best ML/D. Working out exactly what these values are is a little tricky, because the problem is actually to find the the best M and C_L *for the best geometry*. So, ideally you should optimise the operating point and geometry at the same time.

Such a computation is beyond the scope of this unit, but fear not, because the results of these computations have been found and are quite easy to understand. In figure 42 you can see the optimised geometry (i) looks like a conventional supercritical section and (ii) the operating point is for $C_L \approx 0.5 - 0.6$

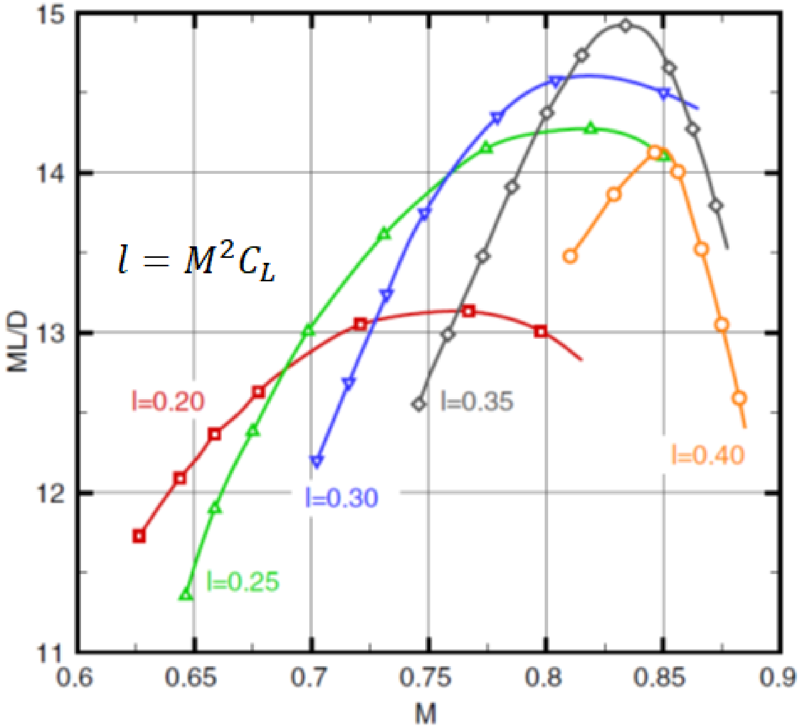


Figure 41: ML/D variation with M for 747-100/200

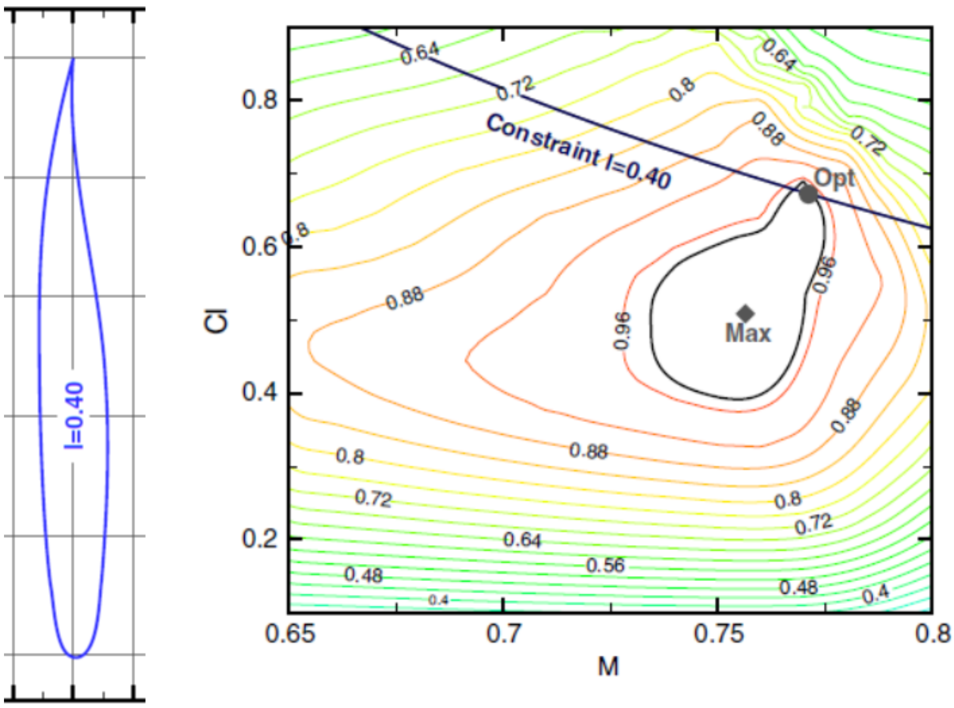


Figure 42: Sketch of ML/D variation with M anc C_L for an aerofoil optimised for ML/D at fixed $l = M^2C_L$

and $M \approx 0.76$. Plenty of assumptions went in to those computations, so don't take them as absolute truth, but they are clearly indicative for aircraft design. Aircraft cruise a little faster than this, around $M = 0.8$, because the computations we ran didn't include wing sweep effects (they were 2D calculations).

8.2 Cruise conditions

Knowing the optimal cruise conditions, we can estimate a parameter like cruise altitude.

Rearranging the lift equation using the speed of sound result $a = \sqrt{\frac{\gamma p}{\rho}}$ gives

$$\frac{W}{S} = C_L \frac{1}{2} \gamma M^2 p_\infty \tag{52}$$

so W/S can be estimated using an approximate $C_{L_{max}}$ of 2.5 at 60m/s (a typical stall speed for a commercial aircraft, and set by runway length requirements)

$$\frac{W}{S} = 2.5 \times \frac{1}{2} \times 1.403 \left(\frac{60}{340}\right)^2 \times 101325 = 563 gkg/m^2 \tag{53}$$

We can also write

$$\frac{p_{\infty_{cruise}}}{p_{\infty_{land}}} = \frac{C_{L_{max}} \frac{1}{2} \gamma M^2_{land}}{C_{L_{cruise}} \frac{1}{2} \gamma M^2_{cruise}} = \frac{2.5 \times 0.176^2}{0.6 \times 0.8^2} = 0.2 \tag{54}$$

We can now look up in standard atmosphere tables to find the altitude that gives this pressure ratio, which turns out to be near 38,000ft.

The explanation of cruise altitude is therefore that this is the altitude for which an aircraft with a wing loading that meets the runway requirements will cruise, if it is powered by a turbofan. Lower, and C_L or M would be below optimal, higher, and C_L or M would be above optimal.

Note that M_{cruise} would be lower for a turboprop ($M = 0.6$ perhaps), increasing the pressure ratio and lowering altitude for optimal cruise. This is why propeller aircraft fly lower.

If you want to recompute for Concorde, we just need the data for optimal C_L at Mach 2, which turns out to be in the vicinity of 0.2 for delta wing planforms, and allow for the higher landing Mach number of 0.22 with a more limited $C_{L_{max}}$ of maybe 1.5 (because Concorde had no high lift devices). This gives a pressure ratio of 0.09, implying an altitude of 55,000ft. Changing these numbers a little won't alter that figure much.

There are some other benefits to higher altitudes, including better thermodyncamic efficiency of engines due to lower external temperature (not above the tropopause though, where temperature is about constant, so this argument falls away), but a primary driver is this compromise between low speed and cruise design requirements, as shown here.

Where would a hypersonic transport cruise? Using the Concorde data at Mach 4 would put it near 85,000ft, or with higher M, up towards 100,000ft. You can estimate other aircraft yourselves (SR-71, X-15, C-172 etc...!).

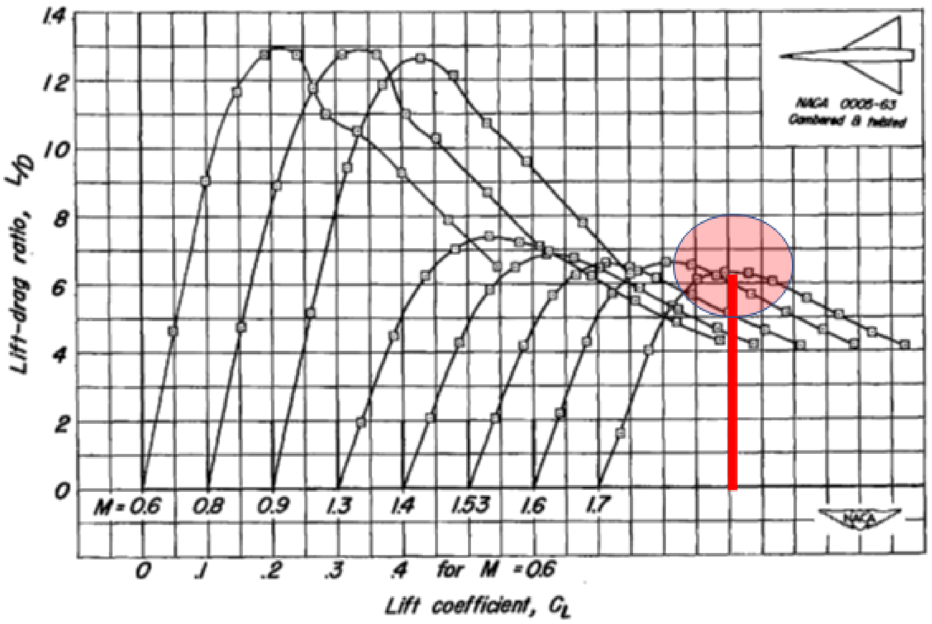


Figure 43: L/D variation with M, plotted for varying C_L, for a delta planform. Red circle shows a peak in L/D for a lift coefficient of about 0.2

8.3 Geometry

We all know roughly what commercial aircraft look like. Long sprawling wings, with smaller tailplanes and tailfins, and some uncomfortable passengers inside. But why is this? Why not make the wings even higher aspect ratio? Or the tailplane for that matter? Why has aspect ratio increased in recent years - did someone learn something new about aerodynamics?

Actually, although we've learnt a lot about aerodynamic analysis in recent decades (CFD), understanding of the underlying aerodynamic behaviour has not altered much. The people who designed the A300, 747, L1011 and DC-10 had about the same overall understanding as those that designed the 787 and A350 - the primary difference was the capability of their numerical models and the availability of new materials, and the availability of supercritical sections. I don't think you could say the same of the 707/DC-8/Comet designers, because a lot of aircraft aerodynamics was learnt during the 60s when transonic tunnels became feasible. However, widebodies of the 70s had already incorporated much of that basic knowledge.

The answer, as ever, comes down to the objective of the aircraft, which we have pencilled in as being Breguet range. You can *always* reduce induced drag by raising the aspect ratio, but doing this increases the weight of the structure, because long narrow wings need more metal to keep them stiff even if their wing area is fixed. Therefore, there is an optimum in aspect ratio, where a maximum in range is found - neither so high the wings are too heavy, nor too low such that induced drag is too high. Designers tread this tightrope to find the optimum.

Wings normally operate near C_L of 0.5 to 0.6 in cruise, where induced drag, which varies with C_L^2 is significant. So, it is worth paying the weight penalty to reduce induced drag by raising aspect ratio.

Tailplanes operate at quite small, often negative, C_L values in cruise, so their induced drag is lower. It is not worth paying the weight penalty for higher aspect ratio, because the induced drag is not that high anyway.

Tailfins normally operate at $C_L = 0$, except for engine-out or crosswind cases. The only reason to have a moderate aspect ratio is to keep the lift gradient up a little, in order to improve directional stability. Tailfins are oversized for normal flight to ensure that in the event of an engine failure at takeoff, the speed does not need to be excessively high in order to retain directional control in yaw.

That's it really. An aircraft's geometry is a weight-aerodynamics compromise.

So why do more recent aircraft have higher aspect ratios?

Simple - composites materials have lowered the weight penalty of higher aspect ratios, so designers have gobbled up the benefit in weight saving to partly increase aspect ratio 'for free'. It's the aerodynamic manifestation of improved materials.

8.4 The right aircraft

There are some interesting anecdotal stories about otherwise similar aircraft. Take for example the 777-300 and 777-300ER. The ER variant increases takeoff weight to allow more fuel for greater range, but the empty weight increases to do this because more metal is needed to take the higher loads.

If you want to fly a longer range flight, only the ER might be able to do this. However, if you wish to fly short routes that don't require the extra fuel weight, the increased empty weight is wasted, and your aircraft will likely be less economical. In other words, you need to match the aircraft to the job.

A similar example is the 747-400 and 747-400D (domestic) variant. The -400 uses winglets to reduce induced drag, as well as having tip extensions, but these all add empty weight, and are not as effective at non-cruise conditions. The wing bending moments are also higher increasing structural weight. Fuel savings are then significant on a long flight, but may not even exist on shorter flights where time spent in cruise is very small. The -400D, intended for domestic flights within Japan, omitted winglets and extensions for this reason (and probably also because Boeing offered a better price on the simpler wings!).

So, it's all about compromise, and understanding *exactly* what your goal is.

A further example is the DC-10/L1011 comparison. Both were trijets of the early 1970s, but where the DC-10 used a 'straight-through' simpler mounting of the middle engine, the L1011 used an aerodynamically more difficult S-duct, placing the engine in the tail cone area.

The DC-10's fin effectiveness was reduced by the presence of the engine, which prevented a taller rudder. In turn, this meant the wing engines were mounted more inboard than on the L1011, because otherwise an engine failure on takeoff would have produced too large a yawing moment. The L1011 gained a reduction in wing bending by having more outboard wing engines (thereby a reduction in structural weight), but had to deal with the S-duct pressure losses on the middle engine.

As it turned out, the DC-10 go to market before the L1011, and therefore outsold it. The tail engine configuration didn't turn out to be a deciding factor, but it's a really interesting engineering example; aircraft design is full of compromises.

Figure 44: Comparison of DC-10 (left) and L1011 (right)

8.5 A small electric aircraft

Electric aircraft are always in the press these days largely due to their good emissions credentials. I will be blunt though - battery electric commercial aircraft are not going long haul any time soon. Why? Well, good range only works if you lower your weight by burning fuel. That flight from London to San Francisco has lost 30% to 40% of it's weight by the time it arrives. If you had to take all that fuel there without burning it, you would quite possibly end up swimming part of the way. Fuel cells do use the mass, so there's scope for optimism there perhaps. Most practical of all is to merely burn hydrogen - jet engines can run on almost anything (just like Diesels - it's only petrol engines that have a fussy diet) - so adapting them to this fuel is feasible.

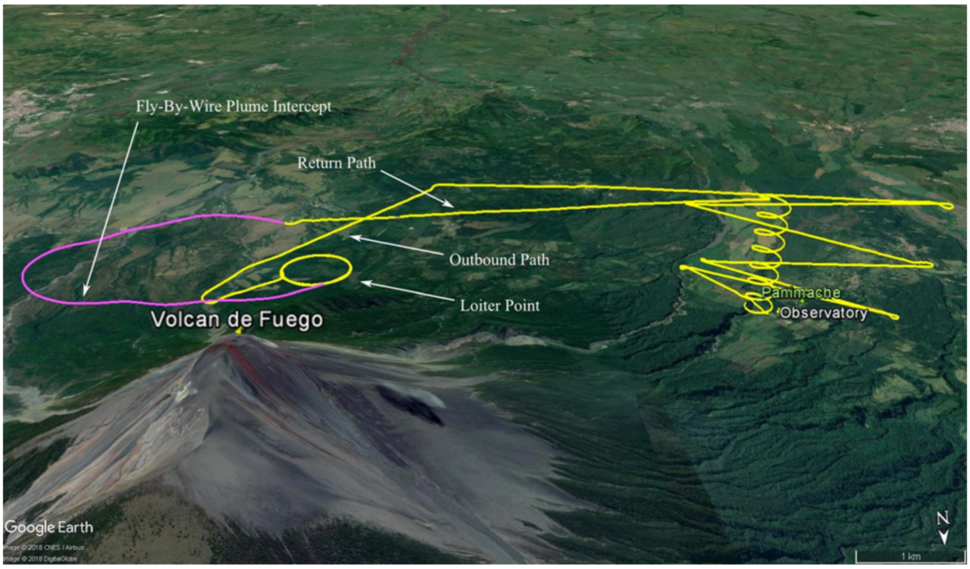


Figure 45: Mission profile for a small electric UAV

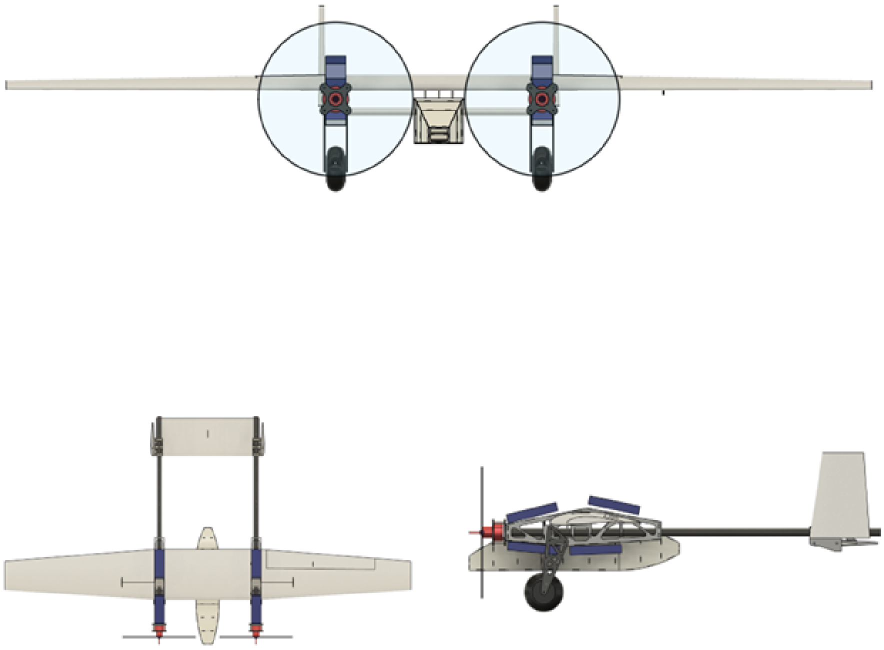


Figure 46: Three-view of a small electric UAV

However, if we we set aside transport category aircraft, battery electric offers lots of benefits for small UAVs. The Breguet analysis is not that helpful in the case where there is no weight change, and the aircraft is propeller powered, so the propulsive efficiency is good and energy usage is power specific. That gives the result that the aircraft should operate at best L/D, or equivalently, for a constrained L then D should be minimised.

$$V = \sqrt{\frac{2L}{C_L \rho S}} \tag{55}$$

$$\frac{d(C_{D_0} + kC_L^2)V^2}{dC_L} = 0 \rightarrow \frac{d\left(\frac{C_{D_0}+kC_L^2}{C_L}\right)}{dC_L} = 0 \tag{56}$$

L/D is

$$\frac{L}{D} = \frac{C_L}{(C_{D_0} + kC_L^2)} \tag{57}$$

Drag for constrained lift and L/D are therefore the reciprocal of each other, so it's the same result either way. Differentiating gives $C_L = \sqrt{\frac{C_{D_0}}{k}}$ and $\frac{L}{D}_{\max} = \frac{1}{2\sqrt{C_{D_0}k}} = \sqrt{\frac{1}{4C_{D_0}k}}$.

If we wanted to minimise power (thereby maximising endurance) then

$$P = C_D \frac{1}{2}\rho V^2 SV = C_D V^3 \frac{1}{2}\rho S \text{ so set } \frac{C_D}{C_L^{3/2}} = f \tag{58}$$

giving

$$\frac{df}{dC_L} = \frac{C_L^{\frac{3}{2}} 2kC_L - C_D \frac{3}{2} C_L^{\frac{1}{2}}}{...} = 0 \tag{59}$$

gives $C_L = \sqrt{\frac{3C_{D0}}{k}}$ and $\frac{L}{D}_{\text{min power}} = \frac{\sqrt{3}}{4\sqrt{C_{D0}k}} = \sqrt{\frac{3}{16C_{D0}k}}$.

So, the minimum power L/D is only $\frac{2\sqrt{3}}{4} = 0.87$ of the best L/D.

The next thing that influences small aircraft is wind, so the aircraft needs to have its optimal cruise speed reasonably high, in this case around 20m/s. This sets the wing loading with an aspect ratio of 8 (chosen as a best estimate) and $C_{D_0} = 0.03$ and $k = \frac{1}{\pi \times 8}$ as

$$\frac{W}{S} = \sqrt{\frac{C_{D_0}}{k}}\frac{1}{2} \times 1.225 \times 20^2 = 21.7gkg/m^2 \tag{60}$$

for a $C_L = 0.87$. For minimum power, $C_L = 1.50$.

Knowing some basic operating points allows exploration of aerofoil options. In this case, as in many others, there is a compromise between C_{D_0} and $C_{L_{max}}$. Good C_{D_0} helps with range, but good $C_{L_{max}}$ is quite important for endurance/loiter because the aircraft has to fly slowly to reach a minimum power condition. If the aerofoil has a poor $C_{L_{max}}$, you may stall before getting to minimum power.

This drives the selection of the aerofoil for this case (figure 47); with a compromise between zero lift drag and maximum lift. Interestingly, after this work was done, a decision was made to focus more on drag than maximum lift, and a subsequent aircraft has moved to using a thinner section. Aircraft design is a moving compromise!

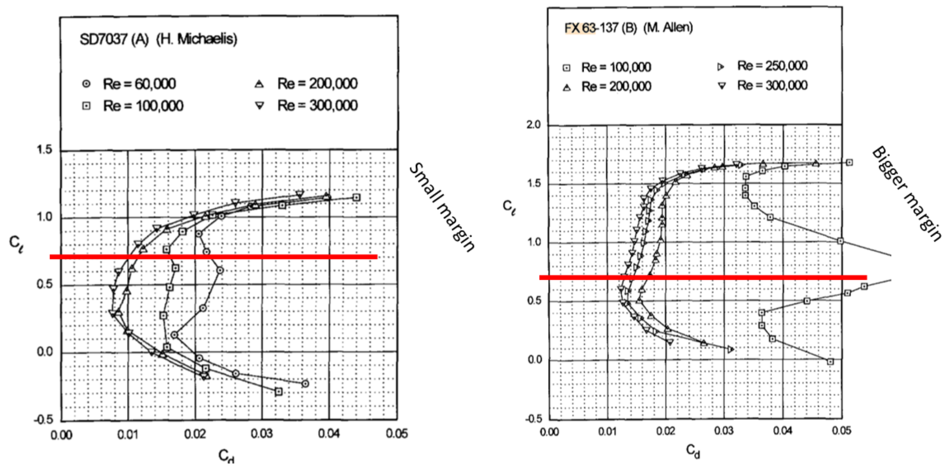


Figure 47: Aerofoil comparison for the small UAV

9 Military aircraft design

I've included this section on 'military aircraft', but to be precise, the arguments contained here are relevant for *all* aircraft, military or civilian, it's just that military aircraft sometimes focus more on manoeuvring performance than their civilian counterparts. However, you will find the arguments presented here apply just as well to a light aerobatic aircraft or manoeuvring search-and-rescue UAV or helicopter as they do to an F-35.

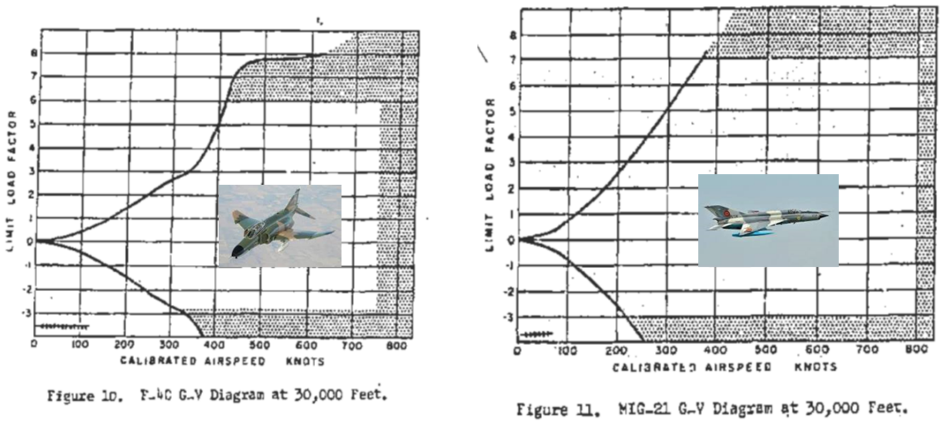


Figure 48: V-g diagrams for the F-4 and MiG-21

The story begins with the somewhat ambiguous comparisons that were made between early combat aircraft. Some had better climb rates, top speeds or rates of turn, and it was hard to draw a scientific comparison between them. This was essentially the state of affairs leading up to 1939. WW2 saw massive advances on all fronts, and in the post-war period a researcher (Christie) and pilot (Boyd) set out to formalise an understanding of aircraft manoeuvring.

WW2 had seen advantage held by an aircraft with the greatest height and speed. This is because height can be swapped for speed, and speed allows an aircraft to manoeuvre. Tight turns, for example, require high lift coefficients that generate very high induced drag, which slows the aeroplane down very quickly. In the dim and distant past I flew a little aerobatics in a light aircraft, and would sometimes pull 5g or so going in to a loop. The aircraft was not fitted with a powerful engine, so a loop required a dive to increase speed, and entering the loop at the bottom of the dive required a firm pull. With a lift coefficient as high as that, and entering a climb, the aeroplane slowed down almost immediately. Similarly, if you attempted a turn at maximum rate, it was impossible to do this for long, because the speed would soon decay. So although 5g+ was feasible, it wasn't feasible for very long - so perhaps not that impressive!

The same was true for WW2 fighters, even with their much more powerful piston engines. Height meant speed, and speed allowed turns. Some aircraft could also add energy more rapidly through propulsion than others, and what was needed was a unified understanding of this for new jet aircraft.

You can see a comparison between two aircraft in terms of flight envelopes in figure 48. Both are limited to about -3g (more than this would be dangerous for the pilot physiologically, with high blood pressure in the brain), while the F-4 has a maximum +6g and the MiG-21 7g. It's worth pointing out that these limits would vary depending on aircraft payload weight. The curved parts on the left of the figure are the stall limits at positive and negative angles, the horizontal lines are structural limits, and the right hand vertical limit comes from a combination of maximum thrust and aeroelastic limits (varies by aircraft).

This figure is very interesting, but mostly it tells us about the structural strength and stall lift coefficient. The aircraft is safe within the envelope, but we have no way of knowing if it *can actually get* to everywhere within it, or stay at that point in the envelope if it does. Some more insight was required.

9.1 Types of manoeuvrability

The first distinction to draw was the difference between *instantaneous* and *sustained* manoeuvring. Instantaneous is what you can do, say, after a dive to pick up speed (like me at the bottom of that loop). The aircraft I flew had excellent instantaneous manoeuvrability. However, with limited engine power once the kinetic energy had been used up, it was a fairly sedate machine, so its sustained manoeuvrability was quite poor (a better engine would have helped!). Figure 48 really tells us about the limits of instantaneous manoeuvrability.

The instantaneous situation is really the height advantage an aircraft might bring to an encounter, ie. it's not linked to the technical design of the aircraft. It's very important, but not something engineering controls, so we will ignore it from now on. Just remember, an Extra 300 *can* outmanoeuvre an F-22, but *not for long!*

Sustained manoeuvrability is much more technical. Christie and Boyd immediately went to energy arguments to resolve this, which is why their work is sometimes known as energy-manoeuvrability 'theory' (although I would contest that it is a theory - to me it seems more of a framework). Semantics aside, the ideas are really important.

9.2 Energy manoeuvrability 'EM-theory'

Aircraft energy is potential plus kinetic

$$E = mgh + \frac{1}{2}mv^2 \tag{61}$$

written per N of weight this is the *specific* energy (SE), with units of length or height often given in ft or m

$$\frac{E}{mg} = h + \frac{v^2}{2g} = E_s \tag{62}$$

we can also define *specific excess power* (SEP) as

$$P_s = \frac{T-D}{W}v \tag{63}$$

where T is thust, D drag and V speed and P_s has units of speed, often quoted in ft/s or kt/s (note that ft/s, the 'climb rate' form, is speed, while kt/s, the 'acceleration' form, is acceleration - see below)

$$\frac{d\left(mgh + \frac{1}{2}mv^2\right)}{dt} = (T-D)v \tag{64}$$

$$mg\dot{h} + mv\dot{v} = (T-D)v \tag{65}$$

$$\dot{h} + \frac{v\dot{v}}{g} = \frac{(T-D)v}{W} \tag{66}$$

A more recent way of writing SEP is as

$$\frac{g\dot{h}}{v} + \dot{v} = \frac{(T-D)}{m} \tag{67}$$

in which case, at constant altitude for $\dot{h} = 0$, it can be thought of just as longitudinal acceleration, or if at constant speed ($\dot{v} = 0$) the climb rate times $\frac{g}{v}$. It seems that SEP is now normally plotted in terms of acceleration rather than climb rate, but both are simply measures of how quickly the aircraft can convert fuel to kinetic or potential energy (and those two are trivially swappable).

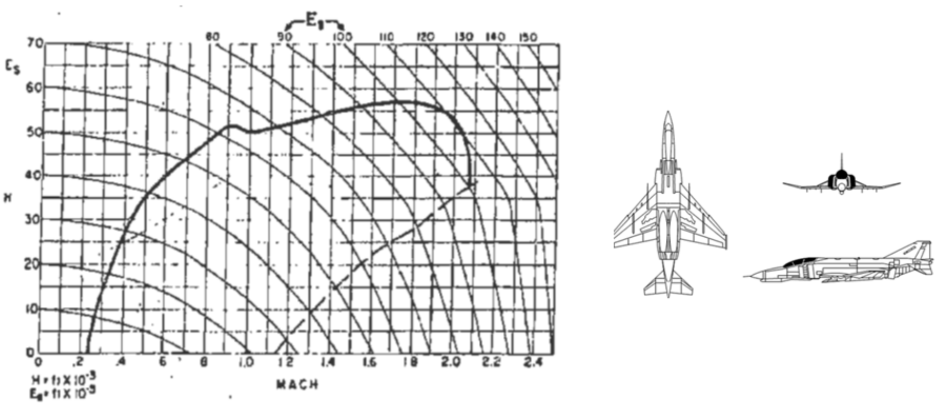


Figure 49: Zero excess power contour for the F-4

We can now look at the contours of specific energy (plotted either in ‘climb rate’ form or ‘acceleration’ form) and the line of $P_s = 0$ for the F-4C aircraft, helpfully plotted by Christie and Boyd in figure 49. In principle, ignoring losses, an aircraft can move along any of the SE contours exchanging height and speed. The right hand side labelled ‘placard limit’ is not a limit with $P_s = 0$, but rather a maximum dynamic pressure limit derived from the aircraft’s structure and controls to ensure safety. The rest of the line (top and left side) corresponds to $P_s = 0$, and the aircraft cannot move beyond this boundary *in equilibrium* - it could however go outside it in an instantaneous manner.

We can quickly see the F-4 can reach a maximum altitude of about 57,000ft at Mach 1.75, or a maximum Mach number of 2.08 at 40,000ft.

The next interesting idea is that we can also plot contours for other values of P_s, as in figure 50. These then correspond to regions within which the aircraft can climb at a rate of at least P_s ft/s in equilibrium. These regions get smaller and smaller as P_s increases, until you reach the 700ft/s contour, which is for around Mach 0.9 and at sea level, and represents best sustained manoeuvrability. Of course, this is merely logical; at low altitude density is higher, so maximum rate fuel can be burnt is greatest, and at higher speed the engine intake consumers a greater mass flow of air. Mach 0.9 is fast, but not for a long slender aircraft like an F-4, and 0.9 is before the transonic drag rise bites. So, 0.9 at sea level is no surprise, and probably many fast jets have their highest SEP contour around here (but who’s contour is the higher? That’s the whole point, of course).

We can now compare two aircraft, the F-4 and the MiG-21, as in figure 51. The F-4 has a higher SEP contour at sea level and M 0.9, whereas the MiG-21 has an interesting SEP contour at Mach 1.7 at 35,000ft. So, each aircraft had an area of relative advantage. But why?

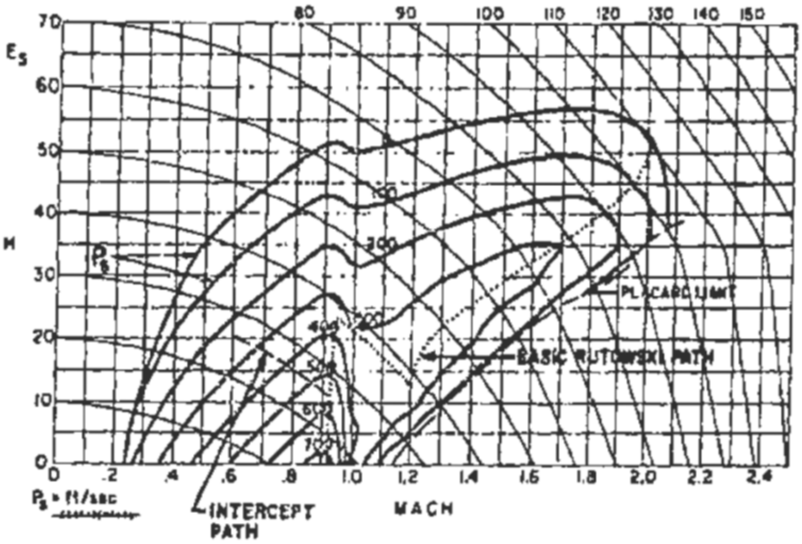


Figure 50: A range of excess power contours for the F-4

Ultimately it's a combination of both T and D. In terms of D, the MiG-21's aspect ratio of 2.2 will provide lower supersonic drag (lower wave drag), while the F-4's aspect ratio of 2.8 will provide better subsonic drag (lower induced drag). There's more to it in terms of the T from the engine, and other aero factors, but that will play a role. Of course, the real truth is that the F-4 was designed as a figher-bomber, needing good range (higher AR), while the MiG-21 was an interceptor, needing high speed and time to altitude (lower AR). Those requirements drove each of the design processes. A similar story plays out in the 5g SEP contours in figure 52.

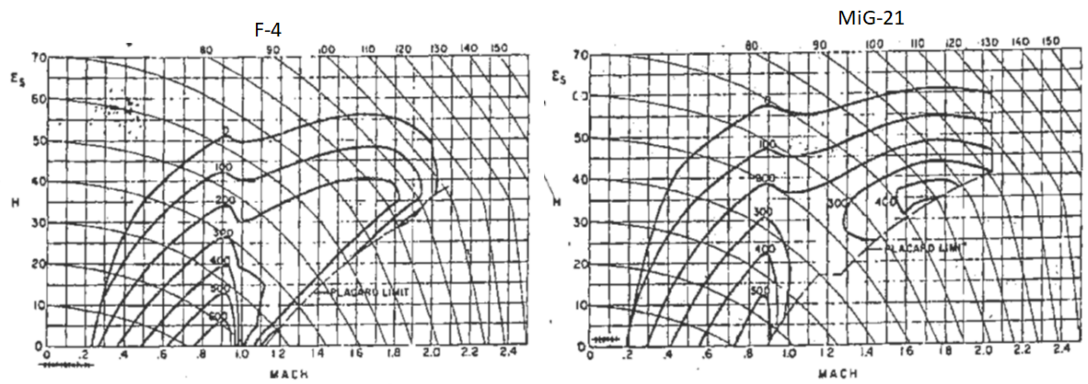


Figure 51: F-4/MiG-21 comparison at 1g

You might also peruse the range plots in figure 53. Interestingly, both aircraft have maximum range at a bit over Mach 0.8 and 35-40,000ft. I would hazard a guess that this is the point that leads to an aircraft C_L in the range 0.5 to 0.6, which is the typical Breguet optimum we discussed before. The F-4 was designed for range, so it's not surprising it has the better range.

If the travails of EM theory are pressing hard on your cranium, let's briefly look at load factor n, which is the maximum acceleration measured in units of g

$$n = \frac{C_{L_{max}} \frac{1}{2} \rho v^2 S}{mg} = \frac{C_{L_{max}} \frac{1}{2} \rho v^2}{\frac{W}{S}} \tag{68}$$

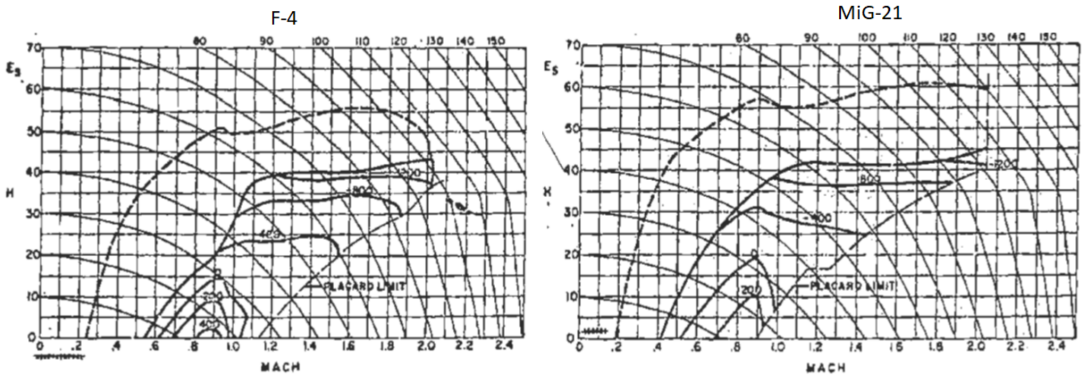


Figure 52: F-4/MiG-21 comparison at 5g

Very clearly, one favours good maximum lift at stall, and a low wing loading W/S. So, wing loading is a critical parameter for manoeuvrability. But think - it represents instantaneous manoeuvrability, not sustained, so it tells a different and complementary part of the story to the earlier EM analysis.

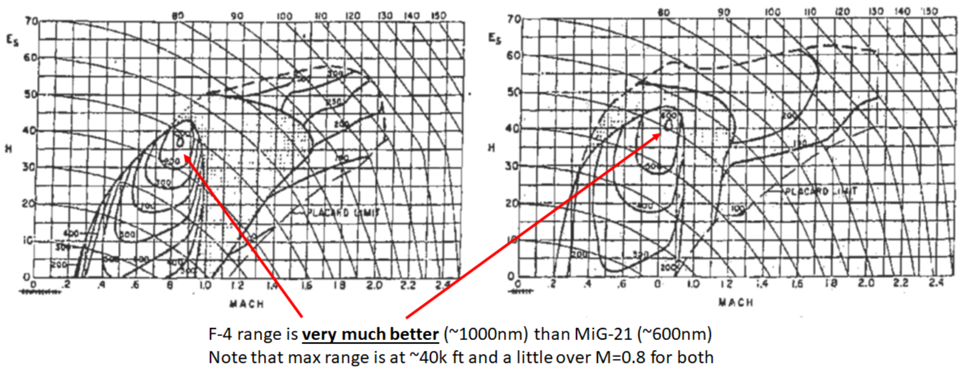


Figure 53: F-4/MiG-21 range comparison

I'm not of your generation, but *Top Gun* was a popular film of the 80s and people of my era often refer to it. If you watch it, you will soon realise that the aerodynamic plot summary is that some aircraft have good instantaneous manoeuvrability, while others have good sustained manoeuvrability. By comparing wing loadings and SEP contours for the A-4 and the F-14, you'll see that what the film depicts makes little aerospace sense. The A-4 out-turns the F-14 at low speeds and altitudes, and has a lower wing loading, but the F-14 has much better SEP contours at high speed. In any encounter, the F-14 would keep its speed/climb rate high and comfortably outmanoeuvre the A-4, while at low speed the situation would reverse.

9.3 Variants of F-35

There's a really interesting comparison in terms of the F-35, an aircraft optimised for performance in different ways, by having three variants, A, B and C. Variant A is capable only of conventional takeoff, and doesn't carry the weight of the lift-fan in variant B, as well as having a smaller wing than variant C. Variant A therefore probably has a higher top speed than either B or C. Variant C has a larger wingspan and wing area, which gives better range through a higher AR and higher fuel loads, but the greater wing area will lower the top speed compared to variant A. Variant C's higher wing area means the approach speeds for carrier landings are lower, and the maximum payload at takeoff is higher. Variant A is by far the cheapest, and the most expensive is variant B, for which the *only* advantage is vertical takeoff, and for which it pays a high price in performance and financial cost. Variant C probably offers better low speed SEP contours, but variant A's will be better at high speed.

So, yet again, the theme of aircraft design being pragmatic compromise plays out - hopefully you're beginning to appreciate the intricacies of this now.

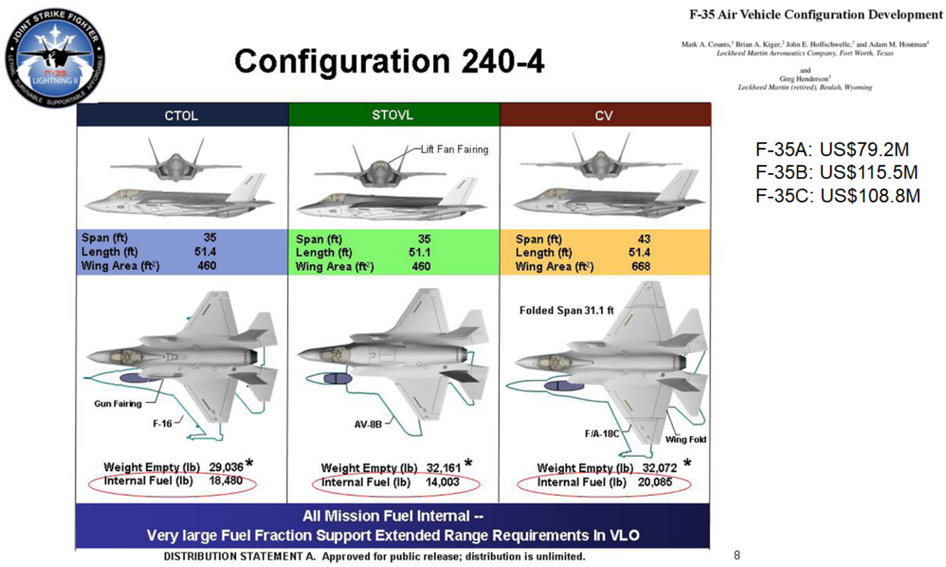
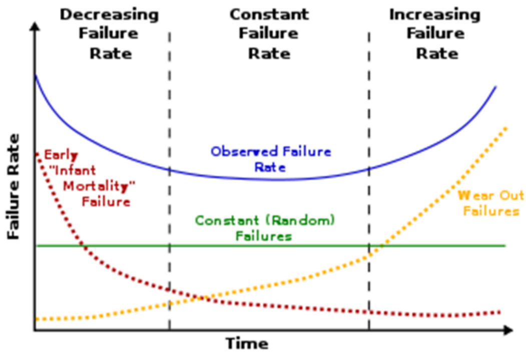


Figure 54: Variants of the F-35

10 System safety

This section is intended to give you some idea about what the internal systems on aircraft do. The glamour of aerodynamics turned my eye a long time ago (so I'm biased), and the methodical process of structural design and optimisation is hard not to appreciate. Perfect spanload distributions and optimised truss ribs paint a beautiful picture of engineering, but there is a great deal more inside an aircraft. It's also really, really important from a human point of view. Only rarely do aerodynamic issues cause fatal problems; normally aero issues are performance related during design. Structural failures have caused many deaths, of course, but structural design has been asymptotic towards a good level of safety for many years now and failures are unusual.

Aircraft systems, on the other hand, are the nasty sting in the tail. If you need three hydraulic systems, as on many commercial aircraft, you can probably guess the sort of earlier tragedies that lead to that requirement. Aircraft systems are a labyrinth of complexity, cunning and contingency planning. From cruise, you can't land a commercial aircraft in less than about 20 minutes even if you're directly over the airport, so you'd better make sure that everyone can survive that long if something breaks. Mid-ocean, it could be several hours before you can land. You need to be able to last the length of a feature length movie while the crew fights to suppress fires, retain electric and hydraulic power and plan a landing. The system reduncancy needs to be planned, and even then it's going to feel longer than the *Strictly* Christmas special and be by far the most frightening thing anyone on board has experienced. System design is therefore about keeping everyone alive, not performance niceties, and only engineers not on board the aircraft and who planned the systems years ago will be able to help them. So, it has to be done properly.



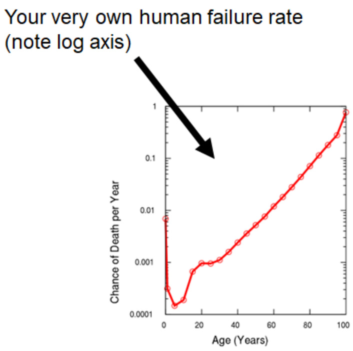


Figure 55: Bathtub curves for component failure probabilities

10.1 Events and acceptable probabilities

Nothing is ever completely safe; there are only acceptable levels of risk/probability for certain events. First, events are usually classified as

- Negligible - little consequence to the operation of the aircraft - eg. an external light fails
- Minor effect - Slight increase in crew workload. Slight reduction in safety margins. Physical effects, but no injury to occupants. A reportable occurrence only - eg. an inertial unit fails, Pitot tube blocks

- Major effect - Significant reduction in safety margins or functional capabilities. Significant increase in crew workload or in conditions impairing crew efficiency. Some injury to occupants - eg. engine failure
- Hazardous effect - Large reduction in safety margins or functional capabilities. Higher workload or physical distress. Serious injury to, or death of, a relatively small proportion of the occupants - eg. engine fire
- Catastrophic effect - All failure conditions which would prevent continued flight and landing. Consequence is a multi-fatal accident and/or loss of the aircraft - eg. structural failure of wing, fuselage or empennage

Alongside this, there are ranked probabilities given by

- Frequent - Likely to occur many times, 1 - 1×10^{-3} per hour
- Occasional - Likely to occur sometimes, 1×10^{-3} - 1×10^{-5} per hour
- Remote - Unlikely, but may possibly occur, 1×10^{-5} - 1×10^{-7} per hour
- Improbable - Very unlikely to occur, 1×10^{-7} - 1×10^{-9} per hour
- Extremely improbable - Almost inconceivable that the event will occur, $\leq 1 \times 10^{-9}$ per hour

To put those probabilities into perspective

- Minor = 1×10^{-3} per hour = Once in 1000 hours. Physical effects, but no injury to occupants. A reportable occurrence only. Once in a lifetime for regular passenger
- Major = 1×10^{-5} per hour = Once 100,000 hours. Significant increase in crew workload or in conditions impairing crew efficiency. Some injury to occupants. Once in career of 3 pilots or life of an individual aircraft (remember aircraft operate for maybe 100,000 hours, while pilots only fly about 25,000 in a career)
- Catastrophic = 1×10^{-9} per hour = Once in a thousand million hours. Consequence is a multi-fatal accident and/or loss of the aircraft. Once in lifetime of the entire fleet of 737 (most numerous civil aircraft - 8000 built, over 1000 flying at any moment). To say a catastrophic event is acceptable at this rate isn't really true, but there must be a probability to use, so this is chosen

Severity	Probability	Analysis
Minor	Reasonably probable	1×10^{-3} per hour
Major	Remote	1×10^{-5} per hour
Hazardous	Extremely remote	1×10^{-7}per hour
Catastrophic	Extremely improbable	1×10^{-9}

Alongside this, types of failure by be categorised too as

- Systematic - These failures will always occur for a given set of conditions. Repeatable and potentially predictable. Software bugs are a good example: once the buggy code is written the potential for the failure is intrinsic to the system and occurrence depends only on the conditions. These are hard to mitigate for and difficult to analyze with rigor. In many cases accidents happen because of events or behavior that were not foreseen. The primary approach is to try and design out this type of fault. Examples: an aircraft skids off the runway due to ice on the tarmac, a computer driving the attitude indicator produces a blank screen if an input data value is greater than expected
- Random - These failures occur during normal operation and are not repeatable or predictable, but can be dealt with analytically using probability. Blowing of light bulbs is an example, or failure from fatigue within design limits. Once extensive testing has determined the probability of failure then 'reliability analysis' can be used to ensure the probability of failure is within acceptable bounds. Examples: a pixel fails on the navigation display after two years of aircraft service, an aircraft skids off the runway because a tire bursts on landing, a computer driving the attitude indicator produces a zero-valued output because a capacitor in the electronics failed

10.2 Redundancy

The probability of event A and event B occurring is the two probabilities multiplied together. Redundancy is the main weapon we use in engineering to reach low probabilities of total failure, and it works really well most of the time.

- Duplex - systems have two lanes. A duplex system can detect faults by cross-comparison between lanes. System operation can only continue after a single fault by pilot selection of the good remaining lane, assuming that this can be identified. This may be done by the pilot
- Dual-duplex - systems have two operating lanes, with two more lanes independently monitoring them. System operation can continue after a single fault, which can be detected and isolated by the monitoring lane. The system can do this automatically
- Triplex systems - have three operating lanes. System operation can continue after a single fault by cross-comparison between all three lanes, and voting out a failed lane. Again this can be automatic. Automatic pilot systems are usually in this mode for making landings in zero visibility

Dissimilar redundancy is a mitigation for 'common mode' systematic failures and can be achieved by using dissimilar hardware/software for each channel, often from different designers or manufacturers. Some examples include having both thrust reversers and brakes for deceleration, or ailerons and spoilers for roll control, or elevators and a trimmable tailplane for pitch control.

Interestingly, the different flight control computers on the A320 use CPUs from different manufacturers, with software from different teams. You don't want to lose an aircraft because there were two copies of the same mistake on board. The elevator-aileron computer uses a different CPU type to the spoiler-elevator computer, and either would be sufficient to land. Dissimilar redundancy is very common in software and processors.

10.3 Probability analysis

We'll briefly consider a statistical analysis to justify our later use of probabilities.

Consider a component that has a rate of failures per hour of λ. This gives a mean time between failures of $\frac{1}{\lambda}$. The failures are randomly distributed in time and independent. This is the situation in the middle of the 'bathtub' curve in figure 55. It is worth noting that this type of distribution is *not* suitable for the left or right ends of the bathtub, because it requires a constant failure rate per unit time. In these regions, a modified distribution would be required.

The probability of no failure in the *small* time interval Δt is then $(1-\lambda\Delta t)$, because the probability of a failure is $\lambda\Delta t$ in that *small* time interval. The probability is $\lambda\Delta t$, because that is how many failures we expect for '1' attempt, ie. $\frac{\lambda\Delta t}{1}$. This is only true if Δt is small (it isn't valid to consider larger Δt until the differential equation below is solved).

So, to find the probability of no failure in the *next* interval Δt, given by $F_n(t+\Delta t)$, we multiply $F_n(t)$ by $(1-\lambda\Delta t)$ to give

$$F_n(t+\Delta t) = F_n(t)(1-\lambda\Delta t) \tag{69}$$

rearranging

$$\frac{F_n(t+\Delta t) - F_n(t)}{\Delta t} = -\lambda F_n(t) \tag{70}$$

taking a limit as Δt vanishes gives

$$\frac{dF_n}{dt} = -\lambda F_n \tag{71}$$

$$F_n = e^{-\lambda t} \tag{72}$$

which is the probability of no failure up to time t (we can check the boundary condition is correct, because for $t = 0$ it is certain there is no failure). The cumulative probability of a failure up to t is then

$$P(0 < T_F \leq t) = 1 - e^{-\lambda t} \tag{73}$$

This can be modified in to a Weibull distribution for more generality, but we will not consider that here. Without getting in to nuts and bolts statistics, the exponential distribution is a good model for independent events occurring at a constant average rate.

The exposure time is usually smaller than the mean time between failures (MTBF) (which is equal to $\frac{1}{\lambda}$), so

$$P(0 < T_F \leq t) = 1 - e^{-\lambda t} \approx 1 - (1 - \lambda t + ...) = \lambda t \tag{74}$$

Essentially, because the failure rate is low per hour, the probability of failure before t is just λt. This is, of course, the same as the $\lambda\Delta t$ result we used above, but we have now justified it for large t providing λ is small, which is, of course, the exact situation for most aerospace components.

10.3.1 Example

Figures 56 and 57 show a system and its failure rates per hour for each component. The overall probability of failure is found using addition of probabilities (where either event A or event B will produce a failure) or multiplication (when both event A and event B need to occur for failure).

An interesting point is that the failure rate per hour increases for a longer flight, despite all components still having the same failure rate. This is because for a longer exposure time, there is a greater probability that multiple failures will overlap and cause a complete failure.

This is, of course, the reason why aircraft on longer flights much have greater redundancy in their systems - for example the ETOPS (extended-range twin-engine operation performance standards). This is a certification standard for twin engine aircraft on flights where the nearest diversion airport may be many hours away (it also applies to three and four engine aircraft, amusingly enough). The ETOPS standards are stricter than normal certification standards to reflect the higher risk of failures overlapping, and the potential severity if they do.

ETOPS is available in different standards, quoted in minutes of one-engine-failed flying time, eg. ETOPS-90, ETOPS-180, ETOPS-240, ETOPS-370 and so on. ETOPS-370 is over 6 hours on a single engine, but I think it would feel more like 6 years for the people on board; it's a long time to look at the ocean with only one engine left. No large aircraft has yet been forced to ditch due in the remote ocean due to system failures, although some have nearly done so after running out of fuel (eg. Air Transat 236). Ditching mid-ocean, potentially in the dark, is of course a more dangerous proposition than in coastal areas or rivers, where rescue would be timely. This drives the fairly strict ETOPS rules.

6 hours flying is over 4 days sailing for a fast ship, so a drop from a rescue aircraft would be needed first. Or, you might get lucky if a ship turned out to be nearby (!). The majority of routes are not ETOPS-370, and the Atlantic is nearly all ETOPS-120 (1.5 days sailing). In any case, the intention is to avoid a remote forced landing.

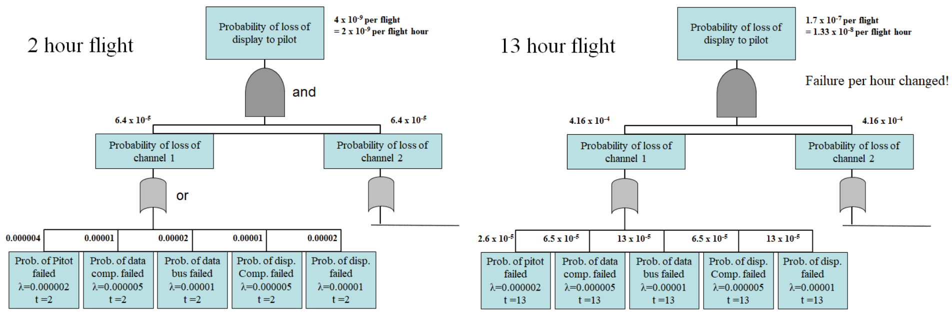


Figure 56: Probabilities of failure for the example 2 and 13 hour flights

Flight length	Rate per hr	Prob per flight		Flight length	Rate per hr	Prob per flight
2	2.00E-06	4.00E-06		13	2.00E-06	2.60E-05
	5.00E-06	1.00E-05			5.00E-06	6.50E-05
	1.00E-05	2.00E-05			1.00E-05	1.30E-04
	5.00E-06	1.00E-05			5.00E-06	6.50E-05
	1.00E-05	2.00E-05			1.00E-05	1.30E-04
	Sys fail per flight	**6.40E-05**			Sys fail per flight	**4.16E-04**
	Both sys fail per flight	**4.10E-09**			Both sys fail per flight	**1.73E-07**
	Both sys fail per hour	**2.05E-09**			Both sys fail per hour	**1.33E-08**

Figure 57: Probabilities of failure for the example 2 and 13 hour flights

10.4 A worst possible flight

Perhaps it's not clear how badly things could go wrong. Afterall, we're so used to flying going to plan that it's hard to imagine what *might* happen once in a billion flying hours. So, what's worst possible set of circumstances that could occur, but which we would still expect to be survivable?

There are plenty of real examples, but let's imagine something that could happen instead. It's easier to think about a fictional event analytically, so here's my attempt. What's written below is composite of several things *that did* happen to actual aircraft - you can read about those yourselves if you wish. The behaviour of the aircraft systems is approximate, because actually each aircraft is quite different in this respect. With that in mind let's join AVDASI flight 999 for a lesson in system redundancy and the unexpected (although I would contest that nothing is unexpected for engineers - we have estimated probabilities for all possible events, afterall, don't we?).

Let's consider a large twin-engine aircraft, flying over the mid-point of a Pacific flight at night. Suddenly, in an 'unprecedented' metallurgical failure, a turbine disk on engine no. 2 fails. Due to the speed and location of turbines, this is not containable, and fragments spew radially outwards through the wing structure and systems, as well as through the fuselage, out of the other side, and in to engine no. 1, destroying engine 1's aceessory gearbox, which drives engine 1's hydraulic pumps and electrical generator. Having fragments travel across the aircraft like this is highly improbable, but it is not impossible, and has happened from time to time. Engine 2 is completely destroyed and a fire begins. The cabin decompresses and the turbine fragments leave a substantial hole in the side of the fuselage, and many passengers and crew are injured by the debris and decompression.

The flightcrew put down their coffee and stow their crosswords. They immediately put on their oxygen masks as the autopilot disengages and the entire cockpit goes dark due to the loss of electrical power. Hydraulic power from the last remaining system drives a motor, which is sufficient to provide electrical power to signal the fly-by-wire controls, but nothing else. The aircraft is put in to an emergency descent, but control becomes difficult because hydraulic pressure has been lost on systems 2 and 3, due to damage from the turbine fragments, and although system 1 remains operational there is no longer a pump driving it after the gearbox was destroyed, and only flying controls powered by system 1 remain functional. Fortunately, this includes the outboard elevator segments and some spoilers. Hydraulic system 1 is completely separate to systems 2 and 3, which prevents it losing its fluid via the leaks in 2 and 3, and accumulators provide pressure until the ram-air turbine is deployed to do so. Cockpit emergency instruments are used, and emergency lighting provides visbility of them. The co-pilot pulls the fire handle for engine 2, which deploys one fire extinguisher and closes the flow of fluids such as fuel and hydraulic fluid. The fire is not extinguished, and the second (last) bottle of extinguisher is needed to finally put the fire out.

The co-pilot ensures cockpit power is switched to the battery, and many instruments now become live again. The co-pilot also uses battery power to start the auxiliary power unit (APU), which then provides hydraulic power to system 1 through its own separate hydraulic pump. The generator on the APU also provides reduced electrical power, and some systems and internal lights come on again.

The aircraft descends below 10,000ft and passengers are able to remove oxygen masks. The flight crew keep theirs on as some smoke entered the cockpit from engine 2's bleed air system, but the main cabin was unaffected because most of it was ducted frome engine 1. The crew switch to bleed air from engine 1, the smoke dissipates, and they remove their masks.

The aircraft is now flying slowly at low altitude, 2 hours away from a landing, with fuel burn increased. The speed is lowered further due to concerns about the size of the hole in the fuselage and the possibility of it tearing further. Aircraft control is degraded because only half the elevators (the outboard sections that were powered by systems 1 and 2) can still be moved and only some remaining spoilers for roll control. The hydraulic actuators for systems 2 and 3 are now dead, but move freely without pressure rather than locking, allowing system 1 to force the actuators of systems 2 and 3.

The situation degrades as fuel is burnt, because some fuel pumps are inoperable due to damage, and the aircraft's centre of gravity starts to creep backwards, making the aircraft more sensitive in pitch. Only manual control is feasible; the autopilot is not suitable for a damaged aircraft.

Weather at the diversion airport is unfortunately bad with strong winds, and without full control in pitch and roll, the approach is very diffcult to fly. The crew are concerned about the extent of damage and begin preparations to land earlier than normal. They extend the undercarrriage successfully, but when extending the flaps, a loud siren blares out to warn them that the flaps are not extending to the same angle on both wings, and the crew rapidly retract them again. Damage to the flap tracks of the right wing from the turbine failure is preventing them extending. The crew leave the flaps up to avoid a fatal roll imbalance and recompute their much higher approach speed for a flapless landing. The crew also decide to use the fuel jettison system to reduce weight for the landing, as they now have approximately 50 tons of excess fuel not required for reaching the diversion airport.

Approaching the airport, engine 1 begins making noises, and an oil warning is notified. The crew decide engine 1 suffered further additional damage beyond the gearbox and has not long left to operate. Due to the higher speed, poor weather and stress, the crew become aware that their approach is too high. They decide to accept the landing rather than go around for another attempt, as engine 1's state is unknown, and if it fails they will not make the airport again. The reduced operable spoilers make it harder to increase the rate of descent to match the correct glideslope, but eventually it is achieved. However, the runway is wet and the autobrake/anti-skid system is inoperable on the right main gear due to wires being severed earlier by debris, so when braking is applied all right main gear tyres lock and burst, reducing braking action and dragging the aircraft right. Further, only some spoilers extend due to earlier hydraulic damage, reducing braking action. The crew lose directional control due to the tyre failures. The aircraft drifts on to the soft grass.

The undercarriage fractures and breaks away backwards, as it was designed to, to avoid damage to the wing structure. Nonetheless, there is a fuel leak and a fire on the right hand side begins. When the aircraft stops on the grass, only the emergency exits on the left side can be used, but fortunately the passengers leave quickly using only half the emergency exits and slides. When the nose gear collapsed the cockpit structure was badly damaged, and the cockpit door completely jammed. The flightcrew open a cockpit side window and leave using a rope, taking their crosswords with them, but leaving their coffee, which has unfortunately spilt.

Things were pretty dicey for a minute there, weren't they? The crew deserve a long holiday, but the team who devised the system architecture did well too. The metallurgical team dropped the ball with the turbine failure, but clever structural design meant the fuselage damage didn't propagate. It was unlucky engine 1's gearbox was destroyed, but the battery and then APU were there. If the APU had failed, the ram air turbine could also have continued to provide hydraulic and electrical power (but that really would have been the final option). It was a pity the flap tracks were destroyed, but lucky there was a warning in the cockpit that the pilots heard and allowed them to quickly put them back in. The fire during the landing was bad, but even half the exits meet the evacuation time. Getting stuck in the cockpit was awkward for the flightcrew, but opening side windows and ropes are handy.

That's the end of our fictional tale, but hopefully it's given some impression of how things can really snowball, and how only a high level of redundancy is going to get you through. The redundancy comes at a weight penalty, but it's worth it.

10.5 Hydraulic

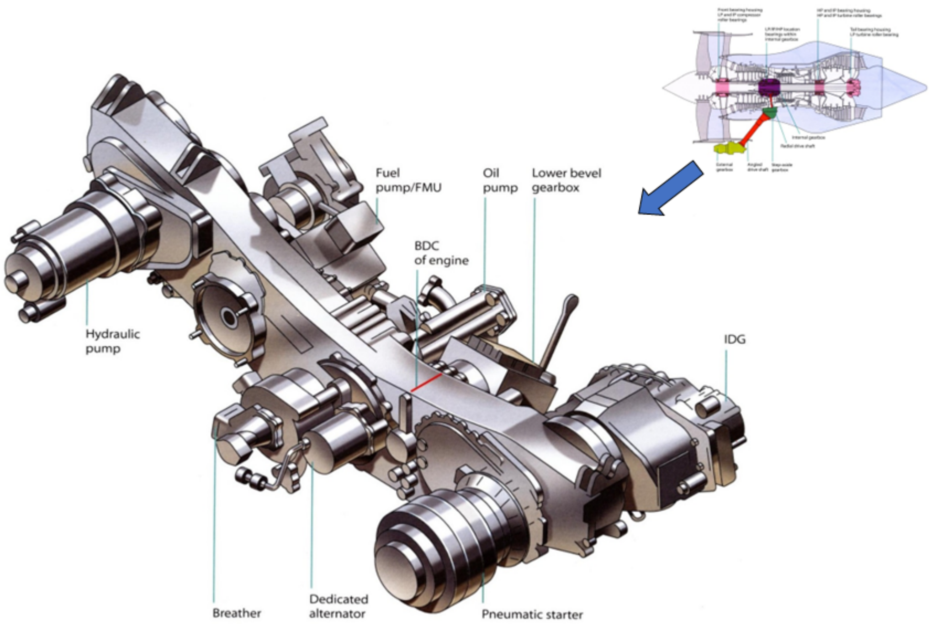


Figure 58: Accessory gearbox

Hydraulic systems are essentially fluid levers. Conservation of energy means a small piston with a small force can be moved through a large distance to produce a large force over a small distance. This is how car braking systems work (together with some pressure assistance generated by the engine). So

$$F_{small}x_{large} = F_{large}x_{small} \tag{75}$$

When you move to larger systems, the input of human power, as in a car brake, becomes so insignificant in proportions to the total force that there is no point in keeping it as part of the system. Instead, power comes from a pump, which runs continuously to ensure the required pressure difference between the *supply* (high pressure) and *return* (lower pressure) pipes is maintained (more on the pump later!).

Hydraulics are great because the system is simple. If the pressure is there, and there is no leak, *the actuator will move*. Compare this wil an electrical motor, which might have power, but still might fail if its bearings are gone, or its circuity blown, or the control wire frayed, etc. Hydraulic action is a level of motive power almost as good as gravity itself.

So, it's natural to want to use hydraulics for flight controls. But, the critical issue, as shown in AVDASI flight 999 is to have multiple systems and pumps, and to prevent fluid loss from all systems simultaneously. It's also important to make sure that a failed or depressurised actuator is still movable by an actuator on another system. Furthermore, in the unlikely case of an actuator completely seizing, control surfaces are segmented to allow the other part of it to continue moving. Left and right

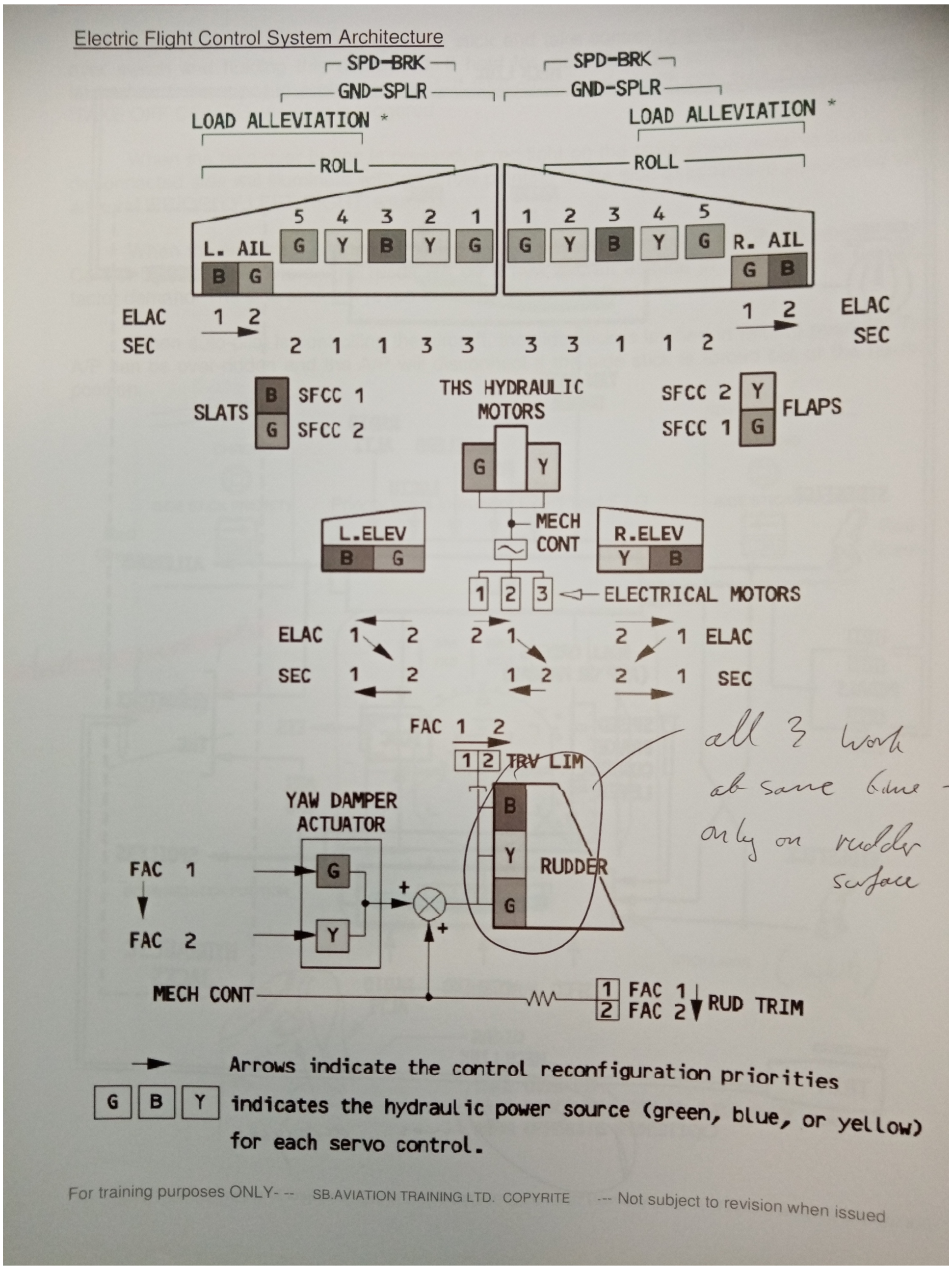


Figure 59: Hydraulic system for A320. the three systems are ‘green’, ‘blue’ and ‘yellow’ or G-B-Y. Only the rudder is powered by all three, because it is the mechanically signalled backup for roll control in the event of electrical failure. Tailplane trim ‘THS’ can also be mechanically signalled

elevators, which are not always segmented, are sometimes linked by shafts, which can be snapped by pilot force alone if one side jams, leaving you with one correctly moving surface.

Figure 59 shows the hydraulic system on the A320 in terms of flight controls. Note how the rudder uses systems all three systems, presumably because it is mechanically signalled and will be the only way of providing lateral control in the event of a complete electrical failure. Also notice how 2 system failures will leave the aircraft flyable, but in a degraded state. If only the yellow system is operable, for example, only one elevator, the rudder, trim and two spoilers on each wing will be moving. Enough to land, but it will not feel normal for the crew.

If you want to read more about the systems on A320, have a look at 'e-rudder', which is a modification that avoids the need for mechanical signalling of the rudder, and thereby saves weight.

10.5.1 Pumps

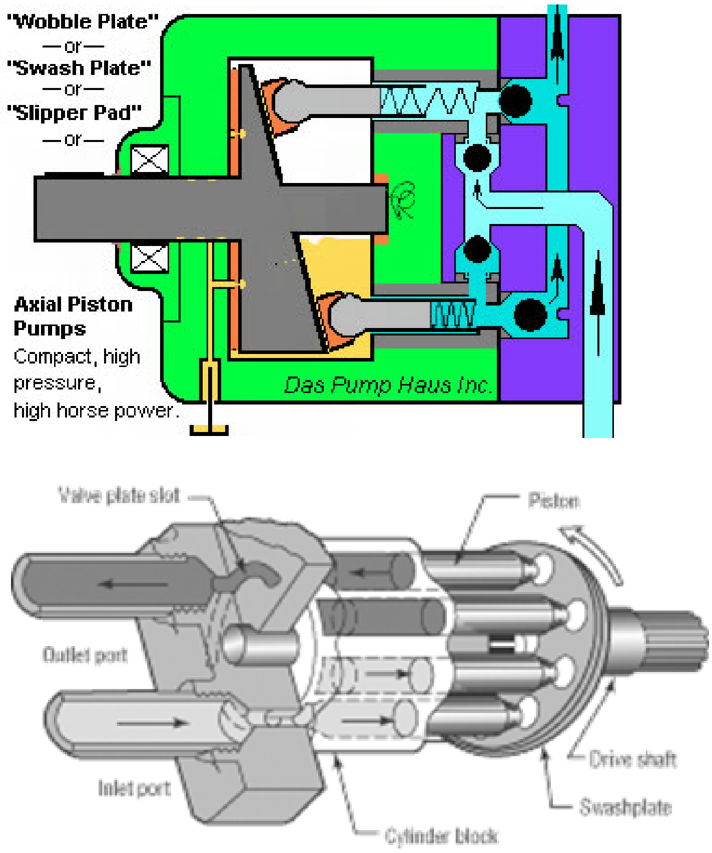


Figure 60: Hydraulic pump and swashplate ('wobble plate')

Ah, the heart of the system.

Auxiliary power generally comes from the *accessory gearbox* on a jet engine. It's driven by a shaft that connects via a gearbox to one of the shafts of the jet engine, and often sits at the bottom of the engine. You will also find generators and other engine pumps attached to the accessory gearbox. It's the place you put anything that needs driving by the main engines.

Hydraulic pumping is variable, however. Sometimes that cheeky pilot is moving the controls around a lot (maybe its windy, or they're just a bit fidgety), and sometimes they are not. So, the pressure of the system needs to be maintained (usually around 3000psi, rising to 5000psi on newer systems) at a constant level, but you *don't know the flow rate*, and it changes all the time.

You could speed up and slow down the main engine to do this, but that would be crazy. The main engine is huge, fairly slow to respond, and has more important things to do, like keep the aeroplane moving forwards. For this reason, hydraulic pumps are super clever, and automatically work out the volume of fluid they need to pump to keep the pressure at the target level.

These pumps use variable displacement pistons, controlled by a swashplate. A series of pistons are arranged in a circular pattern and spin around a shaft, driven by the accessory gearbox. All the pistons connect to the swashplate at one end, and are fixed at the other. If the swashplate is at right angles to the shaft, as the assembly 'rotates', no pumping takes place, but if the swash plate is angled, progressively more and more pumping is achieved. Various non-return valves are used to maintain the low to high pressure flow direction (otherwise it would just go the wrong way!).

In the end the system takes a varying shaft RPM from the main engine, and converts it in to a varying flow rate pump, with the angle of the swashplate controlled to get the pressure just right. Amazing!

Pumps are of course located not just on the main engine accessory gearboxes, but also via a gearbox from the APU (note that this was important for AVDASI flight 999, in the event both other gearboxes are lost).

10.5.2 Accumulators

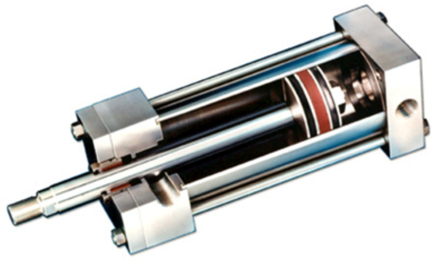
Accumulators store hydraulic pressure and fluid. They range from the huge Victorian accumulators originally used to drive Tower Bridge, where large sliding iron weights were placed on cylindrical tanks and pumped to height, to the spherical and cylindrical ones found on aircraft. Usually these use a membrane or piston with high pressure air on one side, but which is at a lower pressure than the hydraulic system. The hydraulic fluid then flows in to the accumulator until the air is compressed to match the fluid pressure. The device is very similar to the expansion vessels you will find on a central heating system.

If the pump is inoperable then the pressurised air drives fluid out of the accumulator to provide power to the system. However, this is only a short term measure, as the volume of the accumulator is not large.

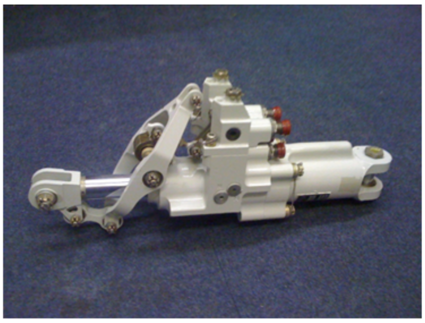
A further benefit of accumulators is to damp out pressure fluctuations in the system, similar to shock arrestors on household plumbing (the ones that can help stop your plumbing from banging). As valves open and close in the system, the accumulator absorbs and expels fluid slightly to dampen the pressure spikes. In the long term, this will reduce wear on other parts of the system.

10.5.3 Actuators

In principle an actuator is just a piston that can have high pressure fluid fed to one side or the other. This is the 'double acting' design. There is a simpler variant which is 'single acting', where fluid only acts on one side of the piston and a spring is used to return the piston, but this is not common in aircraft.



Cut-away of cylinder



Cylinder actuating flight control surface from training jet

Figure 61: Hydraulic actuator

If you pause to think, it's immediately clear it's not at all straightforward. More accurately, an actuator is a 'mini' example of a PID control problem. An input specifies the position of the actuator, and high pressure fluid must be fed to one side of the piston and low pressure fluid removed from the other, until the piston as moved to the specified location.

Early or lower-cost designs use mechanical feedback. The pilot drags a 'spool' (a bit like a plug) to connect the high pressure side the correct way round. As the actuator moves, it drags the spool back in the opposite direction, trying to close the value again. If the pilot doesn't move the control again, eventually the actuator closes the spool itself, because the design position is achieved. The actuator is therefore always 'chasing' the spool. The pressures are very high, so there is little lag in the system, and seems like the actuator always matches desired position exactly.

Note that there is always a 'return' line/pipe, which takes the fluid from the low pressure side back to the system pumps, so that it can be pumped back to the high pressure side, and go round again for another ride.

In a more modern system electronic transducers measure actuator position, and also electrically position the spool. This means any PID behaviour can be programmed, not just the ones that can be mechanically implemented.

It's not really fair to summarise all that in a paragraph. If you want to know more about actuators, go and read about MOOG, a huge and successful company that makes them who have a centre just up the M5 in Tewkesbury. You could spend your whole career working on them (and if you do, you can come back to UoB and rewrite my paragraphs). You may or may not be surprised to learn MOOG sell their actuators to both Boeing and Airbus, and most large aircraft you can think of feature their systems somewhere.

10.6 Electrical

10.6.1 Generators

There's no need to justify why we have electrical power on an aircraft. Super-convenient, with a low-mass distribution system, humans just adore electricity.

So where on the aircraft do we place generators? The same place as everything else that needs power

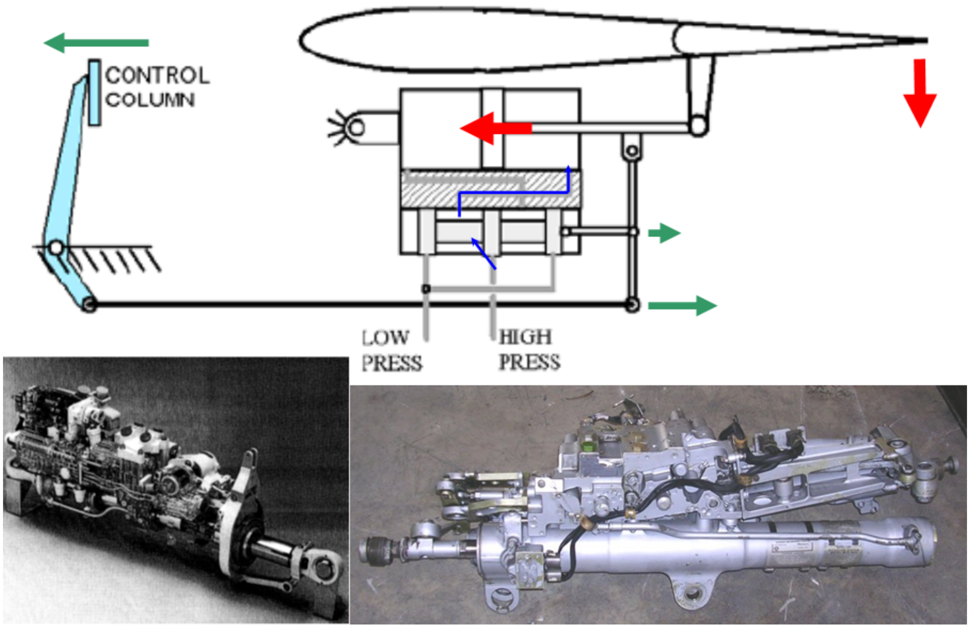


Figure 62: Hydraulic servo mechanism sketch

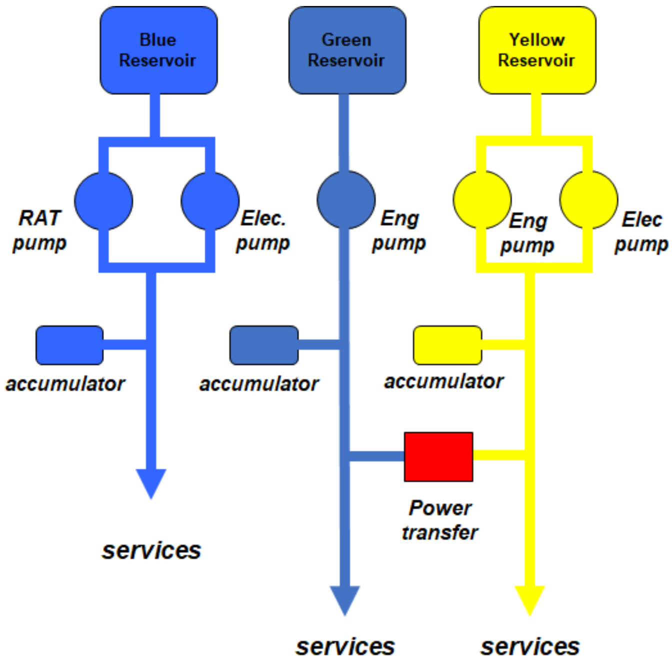
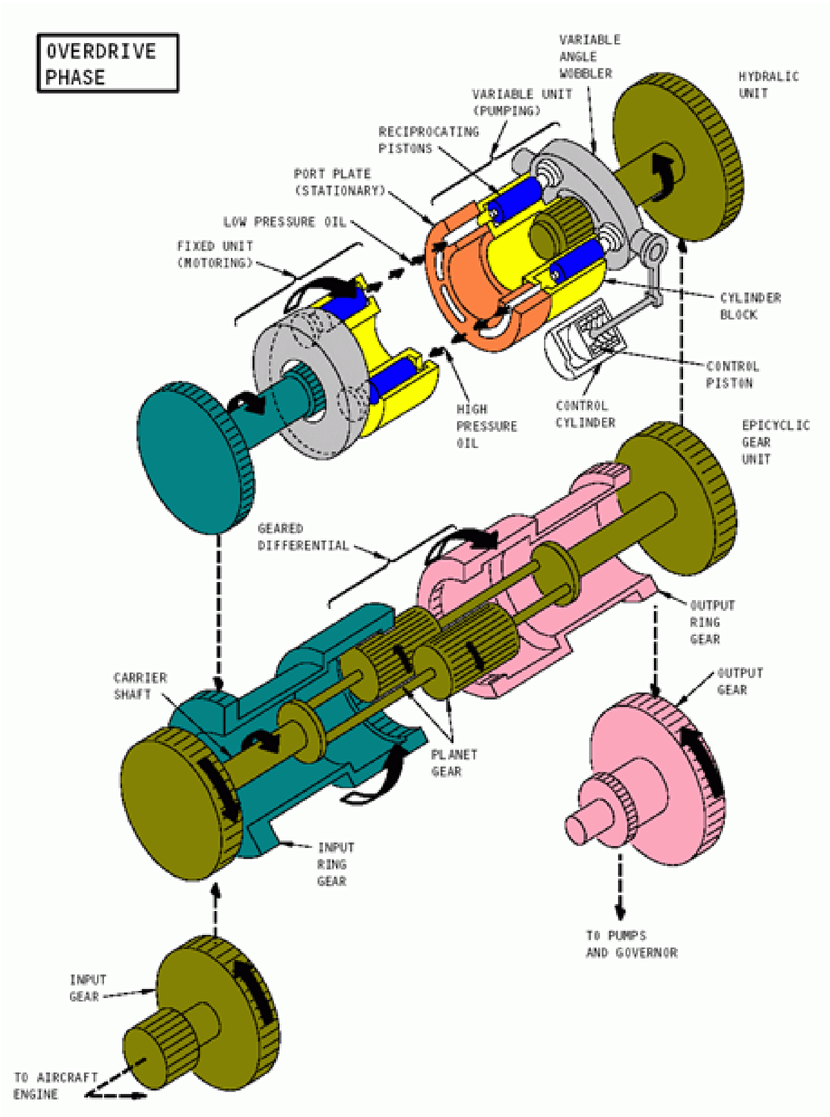


Figure 63: Hydraulic system sketch. Note 'power transfer' - these are usually motor-pumps that allow one system to drive another without any fluid being exchanged



CSD Gear Differential and Hydraulic Unit Schematic

Figure 64: Constant velocity gearbox

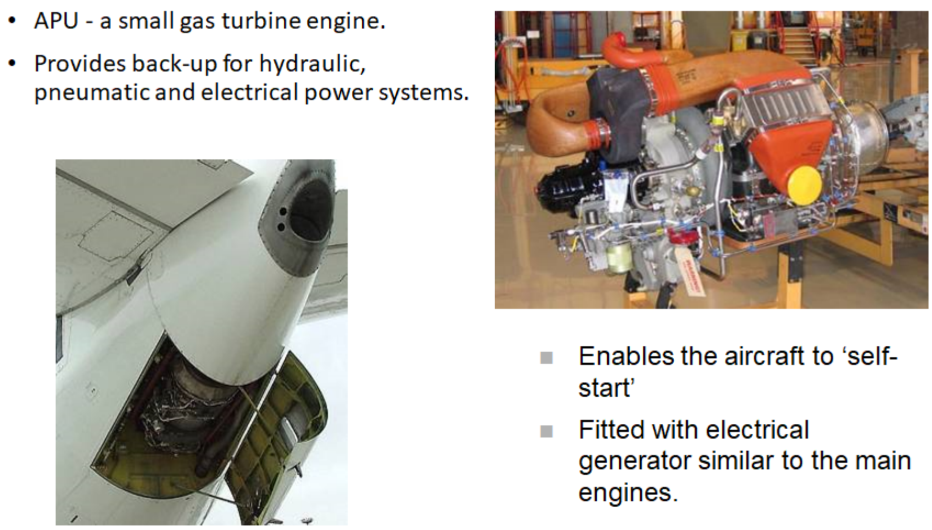


Figure 65: The merits of an APU

Figure 66: Ram-air turbine (RAT) - can provide electrical and hydraulic power form forward speed

from the main engine - on the accessory gearbox. However, we have the same problem with the generator as we did with the electrical pump. Whatever the RPM of the main engine happens to be, we need the generator to spin at a specified RPM to provide alternating current at fixed frequency. We could allow frequency to vary - 'frequency wild' - but this is not suitable for all devices that are connected to the system. So, somehow we need to take an unknown RPM and use it to drive a generator with an unknown load at a fixed RPM.

This is accomplished using a *constant velocity gearbox*, which is a type of continuously variable transmission. The gearbox has two parts (i) an axial differential (same as on a car, but of the type where all three shafts are alined) and (ii) a swashplate motor-pump. The incoming shaft drives a motor-pump arrangement, and the torque that goes from the motor to the pump is controlled by a swashplate. The differential then mixes the two torques together again, and the swashplate is controlled in order to keep the RPM of the system constant nomatter what the load on the generator.

The APU can also drive a generator fitted nearby on most aircraft (see AVDASI flight 999).

10.6.2 Batteries

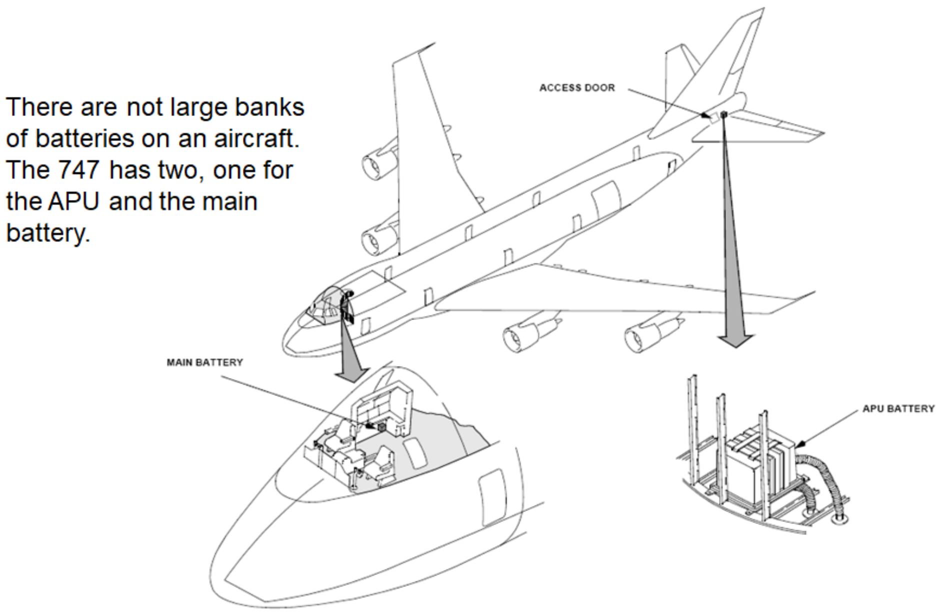


Figure 67: 747 battery locations

Batteries on conventional aircraft are not for propulsion. Their roles are (i) initial power when no ground power is available and (ii) emergency power.

The first area where critical power is needed is obviously the cockpit region. A battery will be located near here to avoid transmission losses on a relatively low Voltage system, and to reduce the risk of aircraft damage severing wires (for example, if an engine fan disk fails, you can expect many wires to be severed, so it's sensible to keep the battery near where it is needed).

If the aircraft has an APU, there will also be a nearby battery for similar reasons (minimise losses and reduce risk). The battery near the APU is capable of starting the APU on its own, after which the APU provides power for the aircraft and starting main engines. There is always a risk that electrical wires may get severed, which drives locating batteries near to where they are required.

10.7 Pneumatic

Why does an aircraft have a pneumatic (air) power system? Can't we just use hydraulics and electricity?

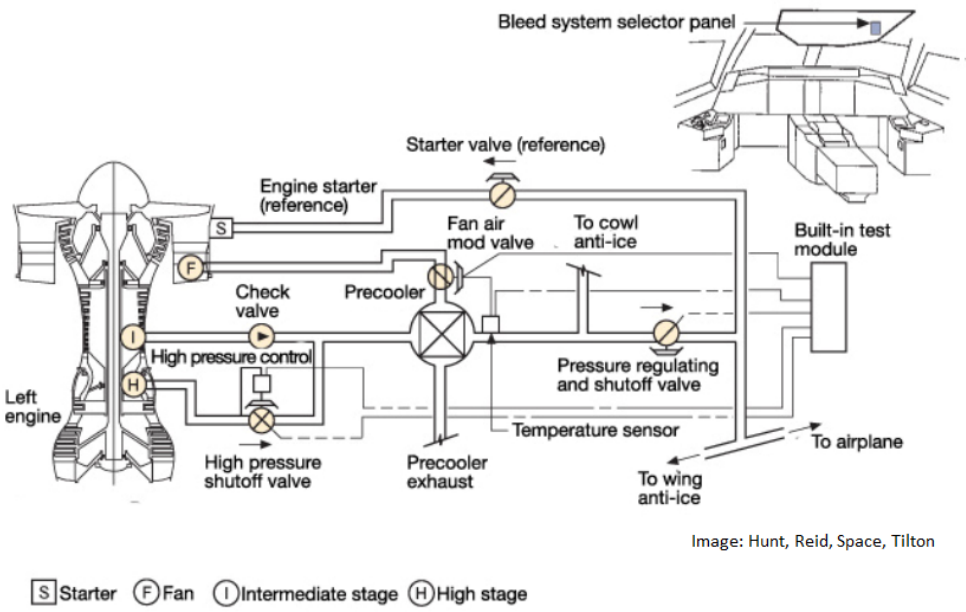


Figure 68: Bleed air system sketch

There are some easy reasons:

- Jet engines product produce large quantities of pressurised air, so it's a convenient and ready source of energy that can be tapped from the compressor stages of the engine, known as 'bleed air'
- The pressurised cabin needs a supply of pressurised air
- Starting jet engines can be done quite conveniently using high pressure air, and a mini turbine or 'starter unit' mounted on the accessory gearbox
- Anti-ice heating can be achieved with bleed air blown through vents in critical areas

Taking pressurised air from the compressor reduces the airflow for combustion, reducing maximum thrust, and effectively adds a load to the engine, reducing fuel economy. It is therefore somewhat undesirable, but nonetheless convenient. The 787 moved to a 'bleedless' design, whereby all the bleed air services were replaced with (usually) electrical alternatives. It was a bold move, and I'm sure other aircraft will follow, but it was quite high risk and lead to increased electrical usage. The A350, interestingly, did not move to a bleedless design. As technology matures bleed air may become less common, but for now it's still a big aspect of aircraft systems.

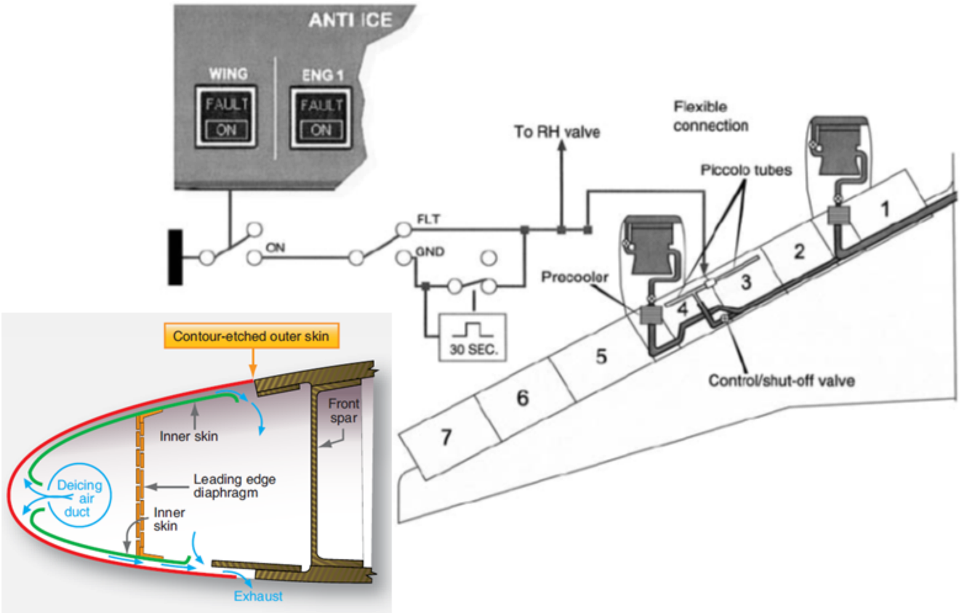


Figure 69: Anti-ice system sketch

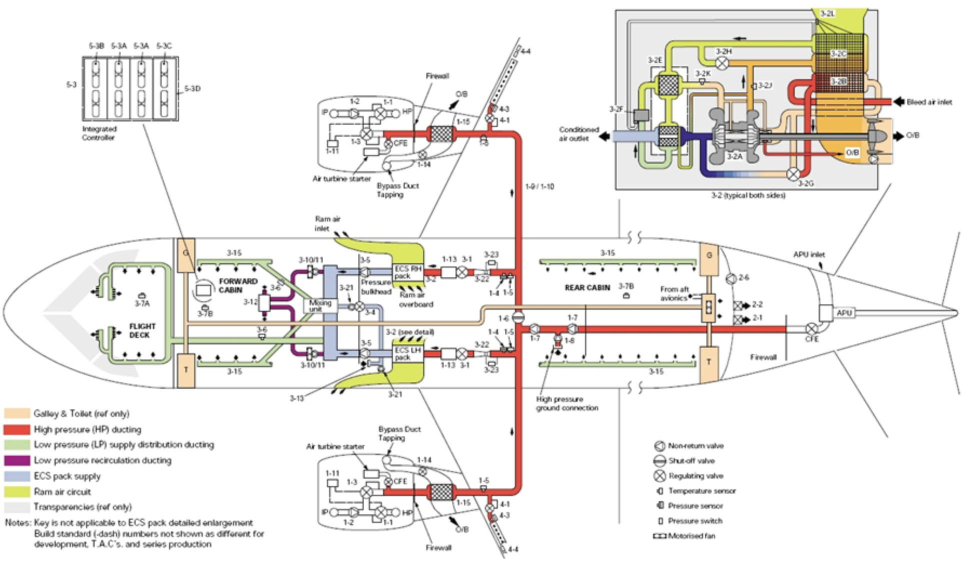


Figure 70: Anti-ice system sketch

10.7.1 Bleed air

Bleed air is taken from one of the compressor stages in an axial compressor design, usually one of the intermediate stages. The pressure needs to be high enough to drive all the required services with a reasonably low flow rate, otherwise the ducts will need to be large, so the pressure does need to be quite high. The air coming off the compressor at this location in the engine is also very hot - several hundred degrees. This makes is suitable for de-icing, but not immediately suitable for cabin pressurisation (unless you like it *really* hot!).

10.8 Cabin pressurisation

The temperature of the air must be dropped to make it suitable for injection to the cabin. Some of this thermal energy is wasted, but some is also recovered by the air conditioning packs, which are *air cycle machines* (ACMs). The amazing thing is that ACMs are actually pieces of turbomachinery, and they are capable of providing cooling air as well we heating. So, the packs can cool you down on the tarmac in hot place, just as well as they can feed the cabin at 30,000ft.

The packs take the air, and cool it through a heat exchanger using ambient air flow (via the intake louvers you see in the belly part of an aircraft, for example). The cooled air is then compressed, which heats it again, followed by another heat exchanger to cool it. Finally the air is expanded over a turbine, which is the one that drives the compressor, before being mixed with a source of the original high pressure hot air to achieve the desired temperature.

So, if air is always blowing in to the cabin, being replaced 25 times an hour, where does it leave? This is done via exit vents or ‘outflow valves’ towards the rear underside of the fuselage. How much they open determines how rapidly cabin air is cycled.

11 Aircraft loads

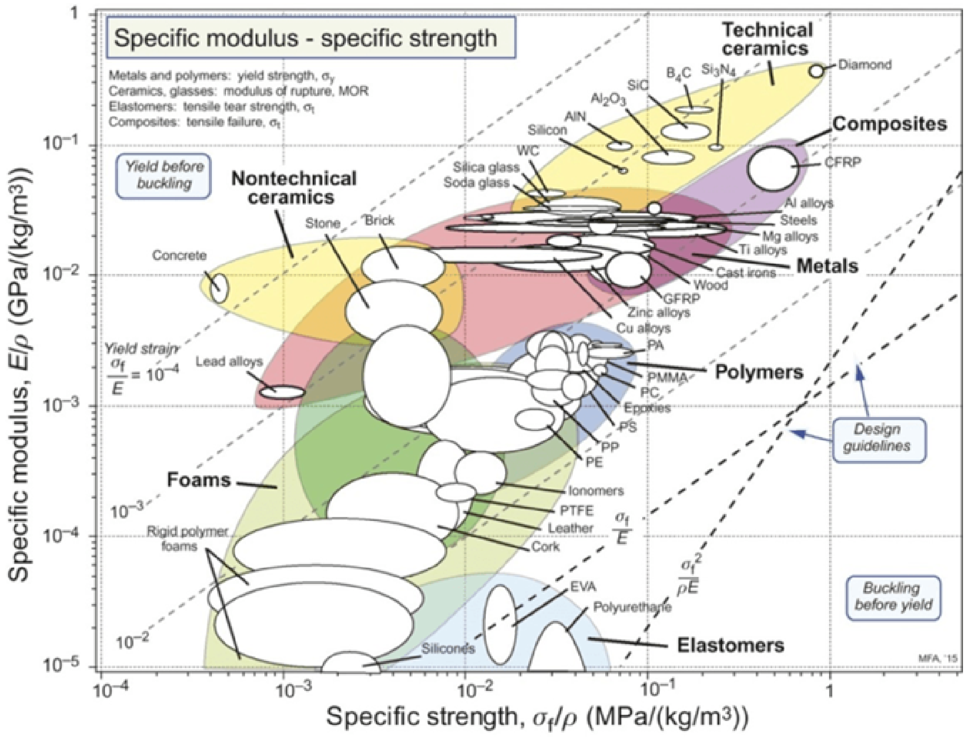


Figure 71: Material properties - the Ashby plot

Aircraft structural design is rather too complicated to cover in detail in a single section, but some points and processes that are a major part of it deserve highlighting.

The first is the standard of safety, or 'safety factor' that aircraft are designed to. By this I mean the ratio between the load which with fail the structure and the maximum likely load experienced in operation. Complicated building structures use a factor of 2, which is not as high as you might think, because in the event of failure other parts of the structure can usually take the loads. Small aircraft have quite high reserve factors compared to the forces experienced in normal flight, especially if designed for aerobatic use.

Large aircraft cannot be designed to very high safety factors because they will rapidly become too heavy. The process, very roughly is

- Find the worst possible load the aircraft will see. This is the worst load for one aircraft over the whole operation of that type, so it is a very rare event. Only a couple of the pilots who ever fly that aircraft type will ever experience it, so it will be an unusual (and unpleasant) experience. This is known as the 'limit load'
- Take the limit load and multiply by 1.5, to give the 'ultimate load'. The structure is designed to fail at ultimate load
- What happens in between limit load, which must be handled without damage, and ultimate load? In this region the structure may be damaged, perhaps beyond repair. Rivets may pop out and certain parts may deform plastically, but no total failure will take place

Figure 72: B-52 tailfin failure

That's about it. Of course, it is intensely hard work to define the limit loads for the aircraft, operating in all potential conditions and in all potential configuration. There may be one limit load with the flaps up and another with the flaps down, and you need to check which part of the structure will actually be at limit load, because not all parts will reach this value at the same time (is it the wing spar or the tailplane spar? The undercarriage leg or the spoiler actuator bracket? All these options will need to be checked).

You can perhaps see why the loads department at a manufacturer is a large department. In comparison, the aerodynamic design group is often quite small! In a large company, the loads are usually defined by one team, and then the structural analysis team will use those loads to identify which parts of the aircraft may be at limit load.

Of course, you cannot evaluate every possibility. Usually, though, in the region of 20,000 load cases will be studied to certify a commercial aircraft. That's hundreds of people doing hundreds of calculations. It's good we now have computers to help.

Don't overestimate the complexity of analysis used. Many of these load cases are handled with the methods you meet in years 1 and 2 for structural design. It isn't possible or sensible to use the most elaborate finite element or CFD methods on every one of 20,000 cases, so very often these use empirical or hand-calculation approaches.

What happens when the aircraft is in service? During design, the loads were estimated, together with the frequency of occurrence, but what about if someone makes a heavy landing and no one is sure if the aircraft is ok to fly again?

Airbus and Boeing (I'm familiar with the Airbus setup, but Boeing will have a similar team) have people on call 24/7/365 to handle this. Data from the event is extracted from the flight data recorder, and sent to the company. They will then use quite rapid tools to see if any limit loads are likely to have been exceeded. If they have been, the aircraft will be grounded until inspections can confirm the state of those components. It costs about £50,000 a day to ground an A320, which is why there is always a team member on call. The final decision rests with the chief engineer for that aircraft type, based on the data computed.

Figure 73: A variety of aircraft loads, some planned by the operator, some not, but *all* expected by the original design team

There is a common anecdote that one of the worst load cases for floor panels arises from passengers wearing stiletto heels, because that generates powerful point loads. It's unusual, but pretty easy to compute, of course.

Briefly, we shall look at an aerodynamic worst case.

11.1 Maximum loads on a lifting surface

A tail fin, similar to that on the A300, has an area of $45m^2$ and can reach a maximum lift coefficient of 1.3. Find the lateral force at stall for the tail fin at $123m/s$ (240knots).

$$F = 1.3 \times 0.5 \times 1.225 \times 123^2 \times 45 = 542kN \tag{76}$$

If we were to assume that this force acted half way up the fin span, at $4.75m$, then the moment is 2574kNm. If this moment were distributed across three pairs of lugs (as on the A300 design) separated by 0.83m (this approximate thickness of the tail fin at the root), then each lug would see a force of 1033kN, which is close to the 905kN failure load quoted by the NTSB in the AA587 accident.

So, working out a maximum force or moment on a surface is fairly straightforward in this manner. The assumption here of course is that the fin reaches maximum C_L. However, the lift coefficient does not exceed this value, so it is a sensible choice.

A similar approach can be used to find maximum loads on wings. A wing can't exceed its stall lift coefficient, so computing loads using this is, ordinarily, conservative. This leads to the concept of a gust penetration speed. This is the speed at which the gust will just stall the wing, at the same time creating a load just equal to the failure load. Therefore, flight in turbulence below this speed is safe, and above it there is a risk that a gust could fail the structure.

12 Beam analysis

It may not have occurred to you, but most major parts of an aircraft are what are commonly referred to as *beams*. Wings, fuselages, tailplanes and tailfins are obvious examples. So, we will consider beams here in a more detail. Keep in mind that there are also *trusses* (like you see on bridges), which are often used for large ribs, and *columns*, which carry large compressive loads - most obviously the undercarriage legs. We won't consider columns and trusses here.

A beam is a structural member subjected to a transverse load, much like the floor beams you may be sitting on right now or the beams in the roof of a house. A wing, for example, sees a continuous load distribution from the lift that is distributed from root to tip all along its span.

When the cross section of a beam is subjected to a moment, an internal moment generating stress distribution must be set up within the beam. In the middle of the beam 'somewhere' there is a *neutral fibre* where the stresses are zero. Above and below this, there is a strain variation that varies linearly with distance from the neutral fibre, which produces a linearly varying stress. The stresses thus produced all have a moment about the neutral fibre, where the moment arm is the distance from the neutral fibre itself.

The strain at a position y above or below the neutral axis is

$$\epsilon = \frac{(R + y - R)\Delta\theta}{R\Delta\theta} = \frac{y}{R} \tag{77}$$

so the stress is

$$\sigma = E\epsilon \tag{78}$$

we know that the applied moment is in equilibrium with the moment of the stree distribution, so

$$M = \int \sigma y dA \tag{79}$$

$$M = \int E\epsilon y dA = \int E\frac{y}{R} y dA = \frac{E}{R}\int y^2 dA = \frac{E}{R}I \tag{80}$$

where I is second moment of area (first moment of area is $\int ydA$, for clarity).

Thus it ever was that

$$\frac{\sigma}{y} = \frac{M}{I} = \frac{E}{R} \tag{81}$$

which is the most important engineering equation of all time. No kidding, it really is - there is no engineer alive who has not used this expression. Careers have been spent applying it and lives lost by forgetting it (or maybe by never knowing it). Unless you leave the engineering field all together, *you will use this expression*. Whether it's bending of pipes, electrical contacts, your arm, toothbrush bristles or aircraft wings, this expression *is* bending. I once used it for a project on inflatable beams, too, which was an interesting twist (no - don't get me started on torsion!).

So, why care so much about it? Well, if you do some extra maths, you can, after a bit of tweaking, integrate $\frac{1}{R}$ along the beam to work out how far the beam will deflect, which is useful. No one likes unexpectedly saggy beams, afterall. Even more useful is that you can work out $\sigma_{max} = \frac{My_{max}}{I}$, which is the maximum stress in the metal. This can tell you if it will fail, or more likely, if local collapse by buckling may take place on the compression side of the beam. It even tells you why 'I' beams are

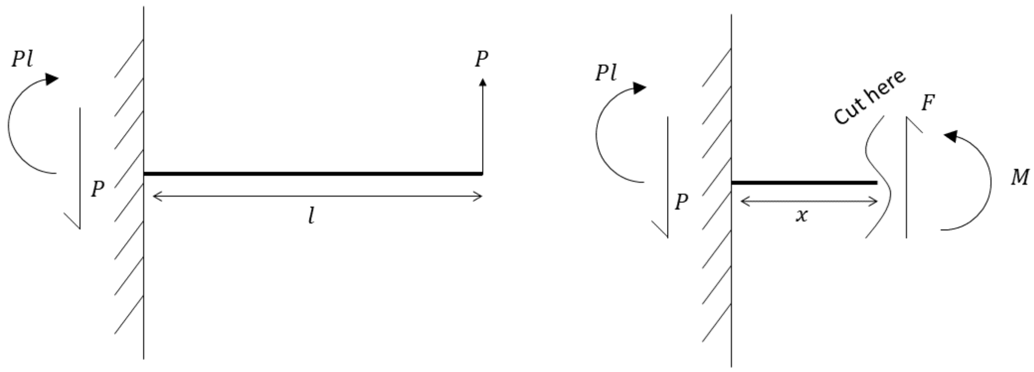


Figure 74: The simple cut cantilever

great. I once used it to find out how far a beam that was holding a car up above my head would bend, and if it was likely to break and drop the car on me. It didn't break.

Of course, none of this counts for anything unless you can calculate M. The good news for all engineering students (who by their nature find long and fiddly calculations irresistibly enticing) is that this is often quite simple to do, using the *method of sections*. It's easy for determinate beams, anyway, which is lots of them. For indeterminate beams you can go and ask a structures lecturer - and take plenty of notepaper with you! No, I'm kidding, indeterminate beams are also quite easy, but require a slightly different procedure that is beyond AVDASI 1.

So, let's begin with our standard cantilever. At the tip of the cantilever is a load P. The bending moment M at the root is therefore Pl, and the shear for F is equal to P.

Next, we chop our cantilever, and ask ourselves the question - what shear force and moment are required on the cut face in order to keep the cut section in equilibrium with the reaction shear force P and root reaction moment Pl? Taking moments about the cut face is a good idea, because the shear force on the cut face then has no moment. However, you can take moments about anywhere you want to.

Balancing moments gives

$$M - Pl + Px = 0 \tag{82}$$

$$M = P(l - x) \tag{83}$$

So the moment is Pl at the root, balancing the room reaction moment, and zero at the tip, where the beam ends. The shear force F is of course P all the way along.

The next important variant uses a distributed load. We could write this out for constant or linear distributions, but let's do a general one and make use of integrals. Then

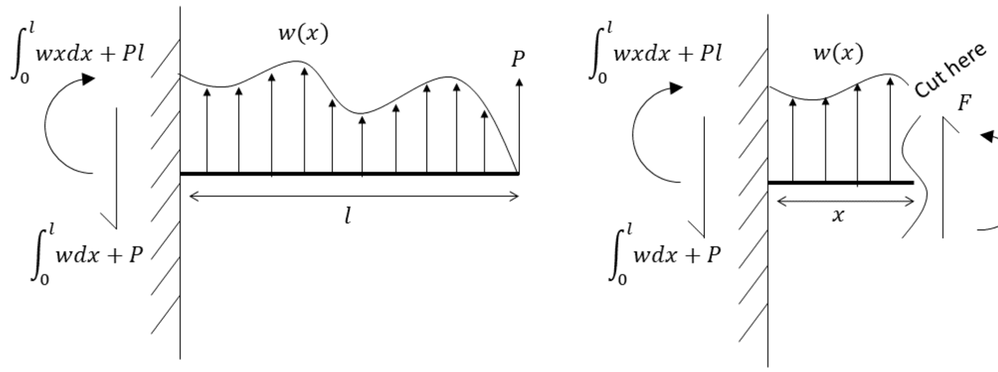


Figure 75: The arbitrarily loaded cut cantilever

$$M - \left(\int_0^l xwdx + Pl\right) + \left(\int_0^l wdx + P\right)x - \int_0^x (x - x') \times w(x')dx' = 0 \tag{84}$$

The equivalent expression for shear force F is

$$F - \int_0^l wdx - P + \int_0^x w(x')dx' = 0 \tag{85}$$

we could choose $w = w_0$ to get

$$M - \left(\int_0^l xw_0dx + Pl\right) + \left(\int_0^l w_0dx + P\right)x - \int_0^x (x - x') \times w_0dx' = 0 \tag{86}$$

$$M - \left(\frac{w_0l^2}{2} + Pl\right) + (w_0l + P)\,x - \frac{w_0x^2}{2} = 0 \tag{87}$$

and

$$F - w_0l - P + wx = 0 \tag{88}$$

but of course any $w(x)$ is acceptable too (eg. an elliptical load along a wing, or even a triangular/linear load). If it is not integrable, the integrals can be implemented numerically. Any arbitrarily loaded determinate cantilever can be handled in this way to determine $M(x)$ and $F(x)$.

If the beam has so many support points that the reaction forces cannot be found from equilibrium (it is *indeterminate*), then enough supports must be removed until the beam is determinate again. Beam deflections are then found at the support points with the reactions as unknowns, and a set of simultaneous equations solved to find the reactions that give zero deflection at the support points. This process requires deflection calculations (covered in your structures lectures), and is beyond the scope for AVDASI 1.

Fundamentals of Aerospace: tutorial sheets and solutions

T. Rendall

March 17, 2023

University of BRISTOL

My suggestion is you work through the problems before looking at the solutions; we can then work through queries during the synchronous sessions.

Historical and other factual information for specific events is included for interest and relevance where it exists; this is of course not directly examinable. Data for any real events has been improvised where relevant and has no guarantee of accuracy (this is merely a question sheet!).

13 Sheet 1

1. The drag coefficient of an 80kg skydiver falling belly first is 1 with a reference area of 0.7m^2, and 0.7 with a reference area of 0.18m^2 for head-first. Find the terminal velocity in each case using ρ =1.225kg m^{-3}.

Ans:

$$v_{term} = \sqrt{\frac{80 \times 9.81}{0.5 \times 1.225 \times 1 \times 0.7}} = 43m/s = 96mph \quad (89)$$

$$v_{term} = \sqrt{\frac{80 \times 9.81}{0.5 \times 1.225 \times 0.7 \times 0.18}} = 101m/s = 225mph \quad (90)$$

Notice that the drag coefficient is not much better head-first, but the reference area is lower. When working with drag on bluff bodies, it is conventional to use projected frontal area as the reference area, and look at C_D (if desired we can easily see the head-first drag coefficient based on 'belly area' is $0.7 \times \frac{0.18}{0.7} = 0.18$). Technically speaking of course, it is $\frac{mg}{C_D \rho A}$ that actually matters for terminal velocity. Speeds up to 300+mph are possible with streamlining and (presumably) strong nerves.

---

2. Hapag-Lloyd 3378, an A310, crash-landed after running out of fuel due to continuation of a flight with the undercarriage extended. Assuming the undercarriage increases C_D by 0.0170 and a wing reference area of 220m^2, calculate the additional power required for flight at 220kts (=113m/s), at an altitude where ρ=1.0 kg m^{-3}.

Ans:

$$P = 0.017 \times 0.5 \times 1 \times 113^3 \times 220 = 2.7MW \quad (91)$$

$$D = 0.017 \times 0.5 \times 1 \times 113^2 \times 220 = 23878N \quad (92)$$

or equivalent to the weight of about 2.4 tonnes. It's not surprising they ran out of fuel. In cruise typical drag normally would be equivalent to the weight of 5-10 tonnes. Note that the aircraft could have been flown more effectively at its best gear-down L/D, which would have been much slower than with gear up. However, the flight management system (FMS) did not include gear drag in performance calculations, which the crew were not completely aware of, although the FMS did meet certification requirements.

---

3. BA308, a 777, crash-landed at Heathrow in January 2008 following a loss of thrust due to ice restricting fuel flow to the engines. Shortly before hitting the ground, the flap setting was reduced from 30 degrees to 25. Why did the crew decide to do this, and why did they not reduce it further?

Ans:

This is an interesting case. Configured to land, drag is high and stall speed low. By reducing the flap angle, L/D was improved, allowing the aircraft to just clear a busy road with a somewhat better glide angle. Reducing the flap angle further would have improved L/D more (best L/D is almost certainly with flaps up when undercarriage is up, although with gear down it may be for a moderate flap angle), but at the cost of increasing stall speed. The speed was already low, so doing this might have precipitated a stall/loss of control. They might also have raised the undercarriage, but this would have taken time and prevented any chance of a normal landing. With much more height, the undercarriage and flaps could have been retracted and the aircraft accelerated towards its best L/D speed - see Air Canada 143 and Air Transat 236, both aircraft that ran out of fuel and made successful glide landings.

---

4. Experiments are planned to measure the drag on model ship hulls. Using the volume displacement of the hull V, speed u, water density ρ and measured dimensional drag R, construct a non-dimensional force coefficient.

   It is now presumed that the drag coefficient of the hull depends on a non-dimensional number that contains g, a reference length l and speed u. Construct a further non-dimensional number Fr using these parameters.

   A full scale liner is 270m long and will operate at $Fr = 0.25$ (12.9m/s). What is the correct speed to measure the wave drag coefficient if a model of the liner is 3m long?

Ans:
The only difference here is that we use $V^{\frac{1}{3}}$ rather than a length explicitly, so

$$C_{D_W} = \frac{R}{\rho u^2 V^{\frac{2}{3}}} \tag{93}$$

Note that in aerodynamics conventionally the factor of $\frac{1}{2}$ is included, but this is not the case throughout marine engineering. It is only matter of convention and makes no difference (except scaling the non-dimensional coefficient by a factor of 2, of course!). The choice of a cube root of volume is again merely convention. Conventions like this are irrelevant, so long as we remember to stick to them consistently.

The second parameter is easily constructed by observing that the square root of gl gives a speed, so

$$Fr = \frac{u}{\sqrt{gl}} \tag{94}$$

This is normally known as the Froude number, after William Froude, who powerfully demonstrated its use in allowing ship wave drag to be found from correctly scaled experiments based on this number (others had realised its importance as early as 1828 for channel flows, so Froude may have had some background ideas stemming from this).

Hey presto, we now know the model must sail at $0.25 \times \sqrt{9.81 \times 3} = 1.36$m/s. If we then measure C_{D_W} from the model, it will be correct for the full scale ship at 12.9m/s.

Of course, we are ignoring things like viscosity and surface tension, and looking purely at gravity-based surface waves here, which is a limitation.

A further footnote is that Brunel is well documented pointing out that ship capacity increases with cube of length, but drag only with the square of length. We can see above that this is nearly entirely true (ignoring any viscous variations linked to Reynolds number), and largely explains the shipping industry's obsession with vast ships (but not so big they don't fit the Suez or Panama canals...). Note that there is no real difference compared to aircraft; you get more volume for your drag on a larger aircraft. Arguments against very large aircraft are based on commercial reasons, not environmental or physics-based reasoning.

---

5. A 1:40 scale powered, sailing and manoeuvring model of a ship capsizes and turns fully inverted in 5.14 minutes (data from report regarding MV Estonia). How long would it take for the same events to take place on the full scale ship?

Ans:
Free-surface effects of ships in motion scale with the Froude number, $Fr = \frac{u}{\sqrt{gl}}$. The Froude number at model and full scale must match. Time is length/speed, and g is the same in both cases, hence time scales with $\frac{40}{\sqrt{40}} = \sqrt{40}$, so the full scale time is $5.14 \times \sqrt{40} = 32.51$ minutes. The interesting point is that this is the same result as for the static Titanic sinking case in lectures; the truth is that if you substitute $t = \frac{l}{u}$ in that working you will discover the Froude number is the same as the non-dimensional sinking time number I defined. A speed and a length define a time, so it is personal preference to use one or the other, although if there is no forward motion it is perhaps easier to use time directly.

---

6. The cabin window on an aircraft fractures completely with area A, leading to a mass flow rate out of the window of $\dot{m}$. Using the parameters $\dot{m}$, A (area), T (cabin temperature), P (cabin pressure) and the gas constant for air R (287J/kgK), construct a non-dimensional mass-flow number M. If the pressure difference across the window is sufficiently high, this number takes the value 0.685. Find the instantaneous mass flow exiting the window if the area is 0.12m^2, temperature 293K and pressure 0.6b.

Ans:

$$M = \frac{\dot{m}\sqrt{RT}}{Ap} = 0.685 = \frac{kgs^{-1}\sqrt{Jkg^{-1}}}{m^2 kgm^{-1}s^{-2}} \tag{95}$$

Remember that a Joule is a unit of energy, and force is mass times acceleration, so a Joule is mass times acceleration times length. Pressure is of course force divided by area.

For the dimensional conditions given, $\dot{m}$=17kg/s, although technically it may only be about 80% of this due to the flow structure around the exit of the window pinching the streamlines and reducing the area a little. If cabin pressure is more than 1.895 times external pressure (and at cruise it is likely 3-4), the flow through the window will actually be sonic at the narrowest point (at a point known as the *vena contracta* just outside the broken window, fortunately not inside). Inside the cabin near the entrance to the broken window the speed would be much lower but obviously still high. See also National 27, DC-10, 1973.

This is one of the reasons Concorde used smaller windows; if one were to break, the bleed air from the engine would be able to keep the cabin pressure higher despite the the lower external pressure. Also, the aircraft would take longer to come down to below 10,000ft, as it cruised at 50-60,000ft. In addition, the aircraft could not be slowed down immediately without an accompanying fuel pumping operation to move the centre of mass forwards again for subsonic flight. Hence, smaller windows to limit the steady state outflow and give time to put masks on. It's worth pointing out that the cabin pressure will eventually stabilise at a pressure where the outflow through the window equals the inflow from the bleed air, and this will be above external pressure by a larger margin if the hole is smaller. Have a read about *time of useful conciousness* and the *Armstrong limit*; at 30,000ft it is about 30s (long enough to put on a mask) but at 60,000ft it is only 5-10s (getting to you mask is going to be iffy), against a likely descent time of a 5-7 minutes (or a few minutes for a normal subsonic commercial aircraft). Hence the desire to limit rate of decompression to give more time to act - and we're probably thinking more about the pilots here rather than the passengers - you need at least one of them to get a mask on or you're in deep trouble.

Interestingly I was once on a 757 when the masks dropped, but no one put them on for a while, because they were too surprised (myself included). That's people I suppose. Fortunately it was only a very minor exceedence of cabin altitude rather than a rapid decompression, so no one passed out.

7. The 2D hinge moment coefficient at maximum deflection and angle of attack for a 30% chord flap is -0.15. The total chord at this point is 1.5m and the flap extends for 1.5m spanwise. Calculate the moment required to deflect the flap, ignoring any 3D effects, at 30m/s, $\rho = 1.225\text{kgm}^{-3}$

Ans:
Remember definition of a 2D hinge moment coefficient as

$$C_m = \frac{M}{\frac{1}{2}\rho V^2 c_{flap}^2} \tag{96}$$

Moment here is per unit span, ie. Nm/m. Hence we need to multiply by the spanwise length to get the full moment as:

$$M = -0.15 \times 0.5 \times 1.225 \times 30^2 \times (1.5 \times 0.3)^2 \times 1.5 = -25Nm \tag{97}$$

These numbers would be typical for a light aircraft on approach. Of course, for ailerons, this would double to allow for both wings. Clearly a degree of mechanical advantage is needed, as well as some aerodynamic balancing, for the controls to be light to the touch. There is some help in that at higher speeds only smaller deflections are needed (or advisable), of course.

8. List advantages of fly-by-wire compared to a mechanically signalled system.

Ans:
Advantages:

(a) Weight saving - so steel wires or pulleys

(b) Control system can be varied across flight conditions very easily and modified with only software changes

(c) Weight saving allows high redundancy to be preserved (eg. backup signal wires)

Disadvantages:

(a) Potentially vulnerable to an electrical failure - this must be mitigated in some way, eg. by manual signalling (A320 (taiplane/rudder), 777 (limited spoilers and tailplane) or a further independent electrical system (often powered by a small generator from one of the hydraulic systems, as on the A350, which allows rudder, elevator and inboard ailerons)

---

14 Sheet 2

1. Briefly describe the steps involved in celestial navigation.

Ans:

(a) Start with approximate position (AP)

(b) Use time, AP and navigational almanacs (time and AP are the inputs to find the right part of the almanac) to find projected locations of chosen celestial bodies on Earth's surface

(c) Measure 'altitude' (angle) of all bodies above horizon

(d) For each body, plot a line along which the angle to the body is equal to the measured angle

(e) For two bodies, locate the intersection of these lines. For three or more bodies, the computed position is the middle of the resulting polygon

---

2. Briefly describe the differences between radio direction finding (RDF), reverse radio direction finding (RRDF), non-directional beacons (NDBs) and VORs.

Ans:
RDF uses a rotating receiver to determine bearing to the beacon, but early equipment was bulky for aircraft. This led to RRDF, where the transmitter beam was swung round from a starting direction at a known starting point, and the receiver would measure the time until peak reception was measured in order to find the bearing. VORs are very similar in principle to RRDF, but no longer require moving parts, and NDBs are very similar to RDF but with greatly improved and miniaturised systems. RRDF itself is no longer used, having given way to VOR. Both NDBs and VORs are being replaced with a reliance on satellite navigation and INS; however, although this will reduce their number they are unlikely to ever disappear.

---

3. Why is a taildragger more likely to bounce on landing?

Ans:
The centre of gravity has to be behind the main wheels, so the acceleration on impact must pitch the aircraft nose up, increasing the angle of attack and thereby the lift, giving the tendency to become airborne again. A tricycle design inherently de-rotates on touchdown, reducing lift. A common technique is to land a taildragger in a '3-pointer' landing, as this reduces tha chance of bouncing. The disadvantage of derotating a tricycle design is that the nosewheel can hit the ground quite hard if not adequately controlled in pitch - see for example flying accidents to the

757, where a hard nosewheel impact could often buckle the upper fuselage or the nose gear area on the lower fuselage.

4. Why do some higher performance gliders take on water ballast for competition flying?

Ans:
At first sight this is completely bizarre - why make a glider heavier intentionally? The issue is resolved why you realise that gliding competitions are often races, so speed matters more than time aloft. The best L/D of the glider is not changed by adding weight, but the speed at which best L/D is reached is increased by adding weight. This means the best glide angle (ie. the point at which you must operate to maximise distance over ground) occurs at a higher speed. In turn, this permits you to overtake other aircraft in the race. The penalty, obviously, is that the rate of sink is increased, but this simply reduces the time aloft and is not part of winning the competition.

The other side to this is thermalling. Clearly, you will gain altitude more slowly if you are sinking more quickly through the air, so carrying ballast makes things worse. This means if thermals are very strong (or there is wave lift, which is massive, and hauls almost anything skyward), you ought to take water because there is lots of free energy available. If thermals are weak, you ought to leave the water behind because there isn't much energy available from the atmosphere and it will slow your climbs, even if your cruise is faster.

Obviously, the water is dumped before landing to avoid undercarriage or structural damage in a heavy landing situation. This makes for impressive imagery of gliders spraying out water at the end of a race. Sometimes heavier aeroplanes are better - who knew?

Why not take water ballast on commercial aircraft? Well, the energy input comes from fuel, which costs, instead of thermals, which are free. Also commercial aircraft select their altitude to ensure they operate close to best ML/D, so extra weight just lowers the best cruise altitude and scales up dimensional lift and drag (lift and drag coefficients for cruise should remain fixed at their optimal values for ML/D). Obviously, that means more thrust and more fuel burn.

5. List the advantages and disadvantages of hydraulic actuation

Ans:
Advantages:

(a) Rapid motion to full deflection
(b) No physical effort required

Disadvantages:

(a) Risk of overloading structure unless a force-feedback loop is included

(b) Vulnerable to loss of fluid or pump - needs careful redundancy

6. List the advantages and disadvantages of servo tabs

Ans:
Advantages:

(a) No hydraulic actuation - simple, low weight system
(b) Natural force feedback to pilot

Disadvantages:

(a) Reduced maximum deflection angles
(b) Rate of motion limited
(c) Controls may need to be locked on the ground to prevent flip-flopping in the wind
(d) Can produce a soft, less responsive feel for the pilot
(e) Possibly requires careful design to avoid oscillatory problems

7. A particular glider obeys $C_D = 0.00725 + 0.0298C_L^2$, a wing loading of 26 gkg/m2 and C_{D_0} of 0.00725. Find the airspeed for minimum rate of descent and the speed for best L/D. The aircraft is now ballasted to increase wing loading to 31gkg/m2. Find the new minimum sink speed and the new speed for best L/D.

Ans:

$$C_{L_{minpower}} = \sqrt{\frac{3C_{D_0}}{k}} = \sqrt{\frac{3 \times 0.00725}{0.0298}} = 0.85 \tag{98}$$

This is the C_L than minimises power, which in this case the the rate of consumption of potential energy, hence this is the condition for minimum descent. Notice that it is quite high; the aircraft would not be far above stall. Minimising power requires slow flight because of the dependence of power on the cube of speed.

$$V = \sqrt{\frac{\frac{W}{S}}{\frac{1}{2}\rho C_L}} = \frac{26 \times 9.81}{0.5 \times 1.225 \times 0.85} = 22m/s = 43kts \tag{99}$$

$$C_{L_{mindrag}} = \sqrt{\frac{C_{D_0}}{k}} = \sqrt{\frac{0.00725}{0.0298}} = 0.49 \tag{100}$$

$$V = \sqrt{\frac{\frac{W}{S}}{\frac{1}{2}\rho C_L}} = \frac{26 \times 9.81}{0.5 \times 1.225 \times 0.49} = 29m/s = 56kts \tag{101}$$

Note that I roughly recreated the aircraft based on an ASK-21. The truth is that although assuming parabolic C_D is ok, it is imperfect. It works very well for induced drag but it is not particularly great as a model for the change in 2D skin friction and pressure drag, which often varies with a lower power of C_L until near stall where it jumps rapidly (this is the other component that must be added to induced drag).

So what's all this about heavier gliders being better? Well, the ballasted glider still has the same best L/D, and the C_L for this condition is still the same (behold - magic of non-dimensional numbers!) but the speed is now

$$V = \sqrt{\frac{\frac{W}{S}}{\frac{1}{2}\rho C_L}} = \frac{31 \times 9.81}{0.5 \times 1.225 \times 0.49} = 32m/s = 62kts \tag{102}$$

which means if your race is a few hours long, you're going to lift the trophy by stretching your lead by 5-6 minutes with every hour, without any penalty on the distance you can cover. So why not take the whole lake with you? Well lets look at the sink condition

$$V = \sqrt{\frac{\frac{W}{S}}{\frac{1}{2}\rho C_L}} = \frac{31 \times 9.81}{0.5 \times 1.225 \times 0.85} = 24m/s = 47kts \tag{103}$$

The glide angle is fixed by L/D, which hasn't changed, so you're sinking more quickly by the same % as your're gaining time. Also, taking the whole lake with you will make ascending in thermals slow.

If it was a race where all the aircraft were just pushed off the top of a tall mountain, you would want to take as much ballast as possible - with the caveat your stall speed would need to remain attainable by your launch (else, erm, no fly, more crash).

It's worth pointing out the heavier aircraft has spent more energy covering the distance, as its greater mass provided more potential energy at the start.

Most of these arguments don't apply in the same way exactly to powered aircraft because their energy comes from fuel and the fuel consumption varies with thrust for a turbofan or turbojet, unless they have inadvertently become gliders. In the fuel exhaustion case, you can choose to glide for best distance or best time aloft (ie. get closer to safety or arrive in the sea later, the choice is yours). On Airbus aircraft the best L/D speed is referenced by a green dot on the speed display, hence 'greed dot speed'. Nonetheless, the idea of a choice between maximising range (hence miles-per-gallon) and maximising endurance is equally applicable even if there is an engine, albeit with slightly different algebra.

---

8. The figure below shows the 'Rutowski path' in boldface red. Explain qualitatively why this represents a minimum time to climb to a particular altitude ('H') and Mach number ('M') or H-M point.

Ans:
The point here is that it is preferable to remain within the highest feasible specific excess power contour. So, for example based on the F-4, one should accelerate as rapidly as possible to

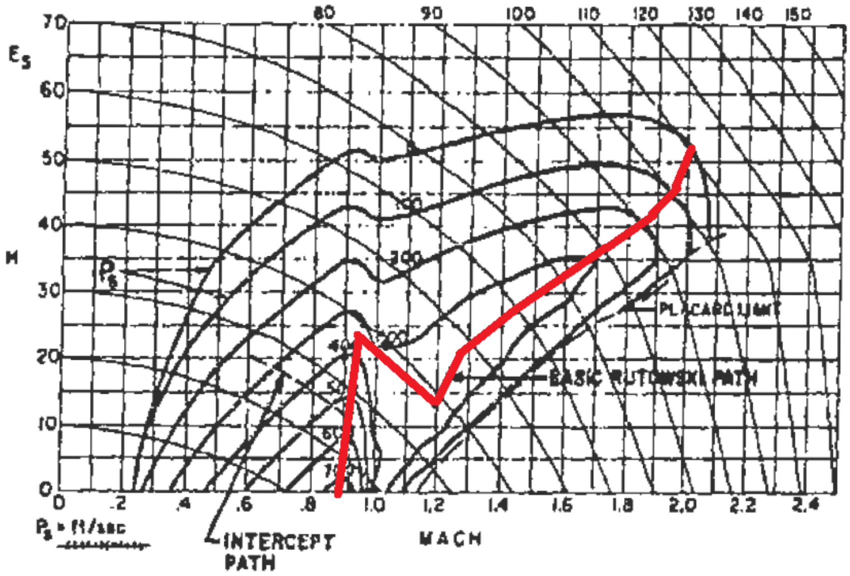


Figure 76: H-M and specfic energy plots for F-4 and A-4

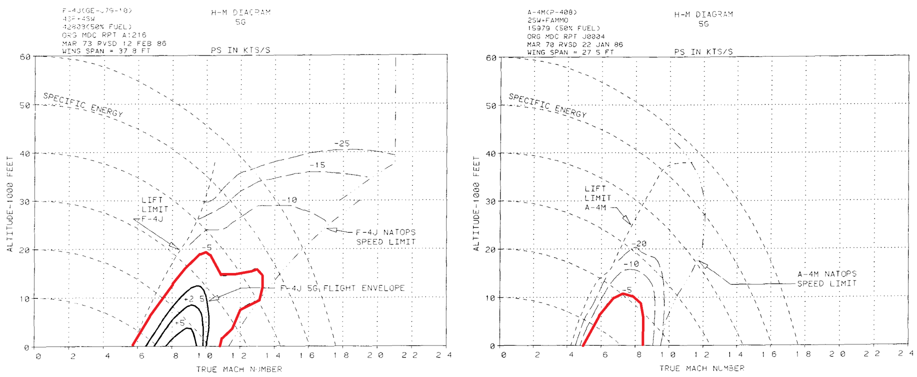
M=0.9, then climb at this fixed Mach number to a little over 20,000ft, descend to a little below 15,000ft while accelerating to Mach 1.2, before climbing again through the middle of the 300ft/s contour (that one that looks like a long pointy finger), then the 200ft/s and 100ft/s contours until reaching the 0ft/s contour and the final maximum H-M location (ie. the highest possible energy state - note that interestingly the highest energy state does not correspond to the highest altitude).

Following this path means you stay within the highest feasible contour at any H-M point. Note that in reality the aircraft is constrained by its ability to manoeuvre (eg. change from ascent to descent, which means the achievable path is a smoothed version of this). Nonetheless, it is true that it is worth sacrificing altitude to gain speed around Mach 1. This is because of high drag coefficients in this region, as well as intake inefficiencies.

Mathematically, the path is formed by connecting the points where the specific excess power contour is tangential to the specfic energy contour. Or, another way, that you look along every energy contour to find the greatest excess power, and use that point. Under this argument, the sacrifice of altitude for speed is just a traversal along a specific energy contour to a new location of maximum excess power.

The same argument drives the use of reheat by Concorde through transonic speed (although whether the objective was minimum time or minimum fuel to altitude is not clear to me); it could similarly be justified by wishing to avoid a region where ML/D was low (this being broadly equivalent to 'miles per gallon'). For a commercial aircraft descending swiftly again from a climb would have been uncomfortable for passengers and their poured beverages, and also with an aircraft of larger mass this transition would take longer, thus negating some of its benefit and probably explaining why there was no descent phase. In comparison, time is critical for interception, and smaller military aircraft can change from climb to descent and back again quickly, much to the amusement of the pilot.

9. Figure 77 shows H-M diagrams for the F-4 and A-4. The A-4 famously appeared in *Top Gun* and the F-4 has performance somewhat comparable but likely slightly inferior to the F-14. Compare and constrast the performance of both aircraft.



(a) F-4 (b) A-4

Figure 77: H-M and specfic energy plots for A-4 and F-4

Ans:

Very obviously the two -5kt/s curves are different sizes, so the F-4 has a wider range of H-M combinations where its excess power exceeds that of the A-4. Note that this is a negative contour, so both aircraft are losing energy at more than 5kt/s outside this boundary at 5g. The A-4 has a region which covers lower altitudes and Mach numbers, and in fact exceeds the F-4 at those conditions.

Let's compare the region at higher altitudes and Mach where the F-4 holds advantage. In these areas it doesn't mean the F-4 would be able to out turn the A-4, because this type of plot makes no assumptions regarding the manoeuvres themselves. For example, the F-4 may not be able to out-turn the A-4 because its wings may start to stall. However, this misses the point; if it is not possible to out-turn, then the F-4 still has the advantage because it could out-accelerate or out-climb the A-4 - either by flying away horizontally or entering a climb. In either case the A-4 would be unable to follow (unless it sacrificed energy at a higher rate, of course).

Remember that the A-4 could potentially start from a higher energy level (height or speed) and use instantaneous (ie. not sustained) manoeuvres that could temporarily be beyond the F-4. However, after sacrificing energy, the A-4 would still end up in the same position. Furthermore, it is hard for the A-4 to escape afterwards.

Now let's compare the region near sea-level and M=0.45 to M=0.5. There is a small pocket here at 5g where the A-4 can operate at -5kt/s but the F-4 cannot. If you were in the A-4, this is where you would want the manoeuvres to be. However, you can surmise that the F-4 would need to make an error to permit this to happen, as it could always move to another H-M point that was more advantageous.

Most of the manoeuvres depicted in *Top Gun* are low altitude and speed - coincidentally the exact conditions where the A-4 would be at its best. I don't have H-M plots for the F-14 and they may be better than the F-4's, which could change the argument slightly.

You can use an H-M argument to explain why a missile is a tricky opponent. Most can achieve Mach 3 or much more using rockets across the altitude range and have vast specific excess power at all conditions - because they're rockets, and their motors don't see an intake of oxygen that is speed dependent. The excess power means they can certainly out-climb and out-accelerate (horizontally) the aircraft, and often can also out-turn because they are small and unmanned with rigid structures, so their 'g' limits are huge (9g is a human limit, missiles can go to many 10s of g). Hence, aircraft rely on trickery using chaff or flares rather than manoeuvres. The only exception to this is if it is possible to out-run the missile and exhaust its fuel supply - this was sometimes done by the SR-71. The best current surface to air missiles are multi-stage and will estimate the aircraft's ability to escape before launch, so if they do launch, it has probably already been calculated that the aircraft cannot evade by running away. Trickery it is then.

10. A fun proportional navigation experiment. Find a friend, because strangers find this unnerving. Tell them to walk across the room. Then, you start from 90 degrees to one side, and walk so that you are always pointing towards them - 'direct pursuit'. You will notice that you drop behind them, and then pursue them from behind (in a nice kind of way, I hope).

    Repeat the experiment, but this time start by aiming at a fixed point (doesn't matter where it is exactly - you may want to place something on the floor to mark the location) in front of them. Hold up your arm and point it towards them before you start, and then adjust your speed (keeping direction fixed) such that your arm remains pointed directly at them at all times (without ever moving your arm). Collision is guaranteed (at the fixed point you chose), so be ready to apologise. Congratulations, you are now a sidewinder missile.

    OK, some of the details are different because a sidewinder keeps speed at maximum and controls direction instead, but you get the idea.

    You can sometimes observe the same principle if you watch a train crossing on a bridge ahead of you above a motorway, or when you're driving up to a 4-way stop in the US. In fact, deflection shooting is the same, if taken from the projectile's perspective.

15 Sheet 3

1. A 'wobble plate' (also 'swash plate') is inclined at an angle θ and rotates with circular speed ω as shown in figure 78. Find an expression for the piston motion as a function of θ and ωt and its radial position r away from the axis of rotation. Then, find expressions for the power and torque required if the supplied pressure is p.

Ans:
Easiest way to do this is solve for the intersection of a straight line with a plane, where the normal vector of the plane is rotating (here about the x-axis). So, remembering the definition of a plane as $(\mathbf{p} - \mathbf{p}_0) \cdot \mathbf{n} = 0$, and that in this case $\mathbf{p}_0$ is the origin, then we just need to construct the normal vector to the plane of the plate and allow for the rotation. The piston is aligned along the x-axis, so we need to find the distance d between the plate and the y-axis, and using figure 78 to determine the rotated normal vector

$$\left(\begin{pmatrix} 0 \\ r \\ 0 \end{pmatrix} + d \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right) \cdot \begin{pmatrix} \cos(\theta) \\ \sin(\theta)\cos(\omega t) \\ -\sin(\theta)\sin(\omega t) \end{pmatrix} = 0 \tag{104}$$

which gives

$$d = -r\tan(\theta)\cos(\omega t) \tag{105}$$

We could also have defined a rotation matrix as

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega t) & \sin(\omega t) \\ 0 & -\sin(\omega t) & \cos(\omega t) \end{pmatrix} \tag{106}$$

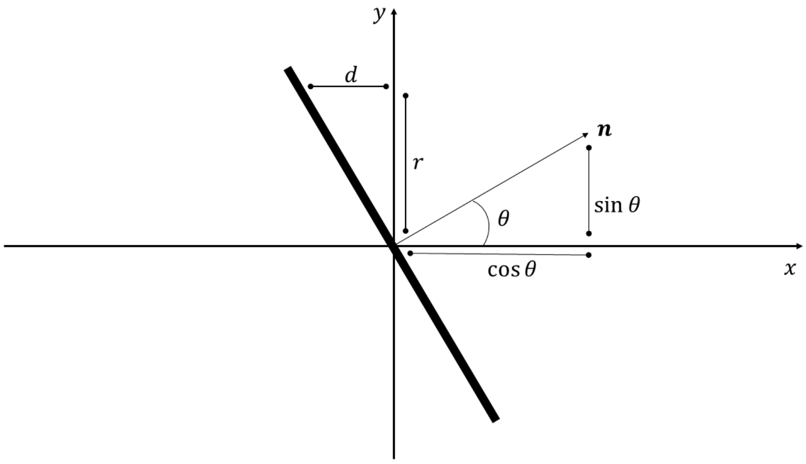
and then found $\mathbf{Rn}$ with $\mathbf{n} = (\cos(\theta), \sin(\theta), 0)$ in order to construct the normal to the plate.

We differentiate to find the speed of the piston, multiply by the cross sectional area A and the pressure p, to give power (pressure times area, times speed) as

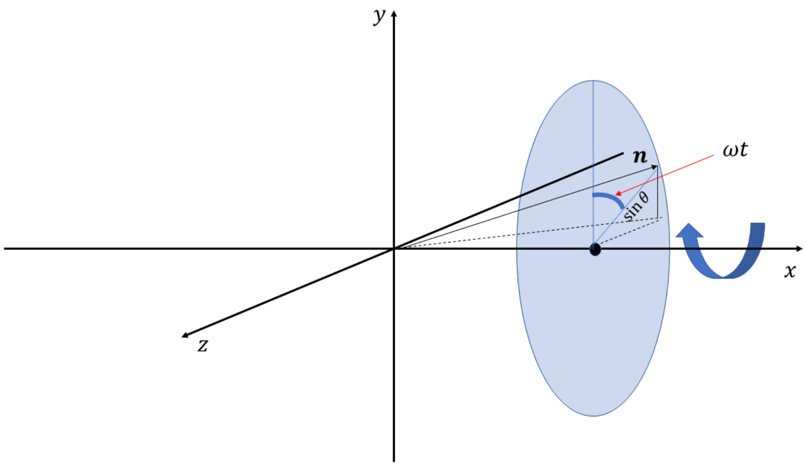
$$P = pA\omega r\tan(\theta)\sin(\omega t) \tag{107}$$

Notice this would integrate to zero over a full cycle; in practice this does not happen, because of the non-return valve on each piston, which means the fluid is drawn back in at low pressure rather than high pressure. Torque may be found from $P = T\omega$, where T is torque.

Notice that this analysis describes both a hydraulic pump, and part of a constant speed gearbox. Also, specific equipment may have different kinematics (ie. between θ and d depending on how pistons are mounted (for example, they may pivot), but for small θ this makes little difference). Have a read about wobble plates (also known as swashplates or slant disks) and their originator, A Michell (also the originator of optimal truss designs!).



(a) 2D Plate angle



(b) 3D Rotation of normal vector

Figure 78: Normal vector and its rotation

2. A small APU for a single-aisle aircraft might provide 100kW. Assuming that this power is delivered via bleed air at 5 bar above atmospheric pressure to a starter unit for the main engines, what is the approximate volume flow rate? Ignore flow speeds and temperature changes (somewhat unrealistic but helpful here)

Ans:
We can use the relation that power is pressure times volume flow rate $P = \dot{V}\Delta p$. This comes from $P = Fv = \Delta pAv = \Delta p\dot{V}$, where we can consider gas at pressure p flowing along a pipe of area A at speed v, giving volume flow rate $\dot{V}$.

$$\dot{V} = \frac{100,000}{500,000} = 0.2m^3/s \tag{108}$$

which is plausible, being about 200 litres a second. Of course, not necessarily all the power goes in to the pneumatic system (electricity might be needed too) and the pressure ratio may be different (eg higher pressure at lower flow rate). This result can also be thought of as coming from the 'steady flow energy equation' that you will meet in thermodynamics.

---

3. The leading edge of a wing is 20m long and the average thickness is 20cm. The aircraft flies through below freezing water droplets at a temperature of -10^o and concentration of 0.5gm^{-3}, with the specific heat of water 4200J/kgK. Assume the water needs to be brought to 5 degrees above freezing to avoid ice formation. Find the power required for the anti-ice system when flying at 70m/s.

Ans:

$$P = 0.2 \times 20 \times 70 \times 0.0005 \times 4200 \times (5 + 10) = 8.8kW. \tag{109}$$

You might assume 2.5 times this if you include both wings and other critical areas (tail surfaces, engine nacelles), so 20-30kW. It's well within what the main engines can provide via bleed air, but remember, aircraft speed could be higher or the water colder (if the water contains no particles to precipitate growth of ice crystals, it can be below zero and still liquid - 'supercooled'). The physics of droplet impact, phase change, splash and runback are all rather more complicated in practice than this estimate, but it gives a rough guide.

It is impossible to protect all parts of the aircraft for continuous flight in the worst icing weather, because water may runback and splash on unprotected areas. This is especially true on aircraft with limited bleed air such as turboprops; see American Eagle Flight 4184 (1994).

---

4. A typical single-aisle aircraft has a fuselage length of 35m and diameter of 4m, and the entire cabin volume is refreshed 15 times per hour (approximate data) with $\rho = 0.96\text{kgm}^{-3}$. Bleed air

enters the air conditioning packs at 150°C, and the cabin runs at 20°C. Calculate the power that must be dissipated by the packs, with specific heat of air at constant pressure $c_p = 1005$J/kgK.

Ans:
We're assuming flow speeds in the system are small. We can then use the 'steady flow energy equation' ignoring velocities.

Flow rate is (note that we haven't worked out the density of the incoming bleed air, but we know the mass flow, because it must be equal to the mass flow exiting to have a steady state)

$$\frac{\pi 4^2}{4} 35 \times \frac{15}{60^2} = 1.83m^3/s \tag{110}$$

$$P = 0.96 \times 1.83 \times 1005 \times (423 - 293) = 230kW \tag{111}$$

To put that in perspective, my old car engine (1.9l 130hp diesel) had an output of 96kW. The air con packs are not merely glorified desktop fans afterall. You can see it is wasteful to lose that level of energy, but the flip side is it is simple (low maintenance), reliable and low-weight (weight is fuel burn, too).

Where does all that energy go? Mostly it is dumped overboard through heat exchangers that are cooled by external air that is ducted over them (creating drag...). Clearly there is a motivation to minimise the bleed air flow rate.

A typical engine on this aircraft might have a total mass flow of 350kg/s, of which about 50-60kg/s would be through the core. Bleed air is therefore small but not insignificant.

The air con packs are actually 'air cycle machines' and they use multiple heat exchangers and turbines/compressors to achieve their cooling. At cruise the turbine/compressor part of the pack may not be operating, but at lower heights and higher external temperatures it will be.

Bleedless designs (eg 787) offer fuel savings, with a slight penalty still in terms of higher electrical loads and weight. A a compressor designed for the cabin air can function more efficiently than one that is designed for the main engine (but hang on - the electrical power still comes from the main engine constant speed gearbox...so it's still the main engine doing it). However, the two manufacturers took different routes - A350 retains bleed air - so the overall compromise (efficiency/cost/maintenance) is evidently closer than a naive 'bleed air is wasteful' argument implies. The more-electric 787 systems have not been trouble-free; in comparison, A350 has seen few similar issues. Perhaps Boeing's more-electric investment will pay off in the *next* aircraft, where their prior experience will help avoid developmental issues.

---

5. Why are batteries often located near the cockpit and APU?

Ans:
Principally there are two reasons (i) it keeps wiring (weight) limited because the battery is near

where the power is needed and (ii) in the event of damage (eg from engine failures, where debris may pass through the fuselage) the risk of interrupted power is lower. Clearly cockpit instruments must continue to function as a priority, and also, an engine failure (perhaps causing fuselage damage) is the exact kind of event where it would be important to switch on the APU as other electrical and hydraulic systems may have failed (due to the engine failure). When in flight, the APU is an important last resort for power, hence the power to start it (the battery) is kept in close proximity. The APU is particularly important on twin engine aircraft, and even more so for twins that fly ETOPS routes (extended twin operations).

6. Based on JAR definitions of likelihood, approximately how often should a major failure occur?

Ans:

- Once in the career of one pilot
- **Once in the career of three pilots**
- Once in the lifetime of a regular passenger
- Once in the lifetime of three regular passengers
- Once in the lifetime of a fleet of aircraft

7. When designing aircraft systems, which approach can mitigate for systematic errors?

Ans:

- Introducing redundancy
- **Using dissimilar hardware**
- Using real-time processing
- Making a system deterministic
- Keeping a record of all design data

8. Fly by wire means:

Ans:

- Flight control surfaces actuated by wire cable and pulleys, as opposed to a push rod
- **Having no mechanical link between the pilot and control surface**
- The use of electrical autopilot to fly the plane
- Power for flight surface actuation provided cables
- A round the pole aircraft

9. On a conventional civil aircraft, which of the following functions is not provided by the pneumatic system?

Ans:

- Anti-icing
- Air-conditioning
- Cabin pressurisation
- Engine start
- **Emergency power**

10. Hydraulic systems are commonly used for flight surface actuation because:

Ans:

- **They can be easily assembled in redundant architectures**
- They are faster acting that other actuation technologies
- Only hydraulics can operate at the low temperatures of cruising at altitude
- They are easy to maintain

11. Batteries on a civil airliner provide power for:

Ans:

- Starting the main engines
- **Starting APU/emergency back-up**
- Electronics that need a 28V DC supply
- The galley
- Electric functions when on the ground

12. A system on an aircraft consists of n components, each with probability of failure in one hour of P. Two systems operate in duplex, so that both must fail for an overall failure to occur. Show that the total failure probability (ie. both systems) per hour is n^2P^2t where t is flight length.

Ans:
If component 1 or component 2 or component n fails, the system fails, so probability of one system failing in a flight per hour of flight is (remembering there are 2 systems)

$$p_{sysfail} = 2\frac{nPt}{t} = 2nP \tag{112}$$

which is independent of flight length. However, for an overall failure, both systems must fail, giving

$$p_{totfail} = \frac{(nPt)^2}{t} = n^2P^2t \tag{113}$$

so the probability is now proportional to flight length.

Why? Well, the probability of dual failure over the whole flight is increasing with the square of time, whereas a single failure is proportional just to time. That cancels to linear per time and constant in time when expressed per hour. Per hour, there's just a higher chance of multiple failures overlapping at some point in the flight if you spend longer in the sky; you can in fact argue this is a reason why aircraft for long over water flights (ETOPS) have to show a greater level of reliability and have stricter certification requirements.

Rest assured that $n^2P^2t^2$ is much smaller than nPt because P is very small. Obviously, it's better to have small n (not too many critical components for each system) and small P (reliable parts). Further redundancy increase the problem to cubed probabilities, which are even smaller, but then total weight will increase. Eventually further redundancy is useless, because there are other things that could happen that are worse and more likely.

From a practical point of view, having more redundancy reduces the probability of a total failure, but it *increases* the chance of a system failure, because there are more systems. This just represents the higher a maintenance cost of multiply redundant systems (but the benefit is many fewer serious accidents!). The aircraft with 10 redundant systems 'won't crash', but there will always be 'something' wrong with it. In some ways this is handled by 'minimum equipment lists', which allow aircraft to fly with known failures providing other systems remain working.